
万卷方法
STATISTICS FOR PEOPLE WHO(THINK THEY)HATE STATISTICS
爱上统计学
Excel 2010 Edition
（美）尼尔·J.萨尔金德 著
史玲玲 译
Ice cream Sales Per Season
Summer
Fall
Winter
Spring
重庆大学出版社

《爱上统计学》读者推荐

诚如其名

非常好的一本书！在此之前，因为要做调查分析，已经自己购买和借阅了无数本类似的书，可是读后总是感觉盲人摸象，而这本书就像是最后一个烧饼，让我觉得自己一下子醍醐灌顶、豁然开朗。这本书非常清晰地阐明了整个抽样调查、统计检验的思想和逻辑，特别是书中总结的一个表：在什么时候需要用到什么样的统计技术，对刚入门的人特别有用。比如什么时候用独立样本T检验，什么时候用非独立样本T检验等。

当然这只是一本入门书，书中对因子分析、回归分析等比较复杂的统计检验方法介绍得比较简略，如果想了解这些的，得另外找书来看。

强烈推荐给还没有入门的人！

——当当读者wuly评

这是一本没有被老师选中的好教材

我从来没有阅读过这样的数学教材。

作为一个大学生，学了十多年的数学，看到的数学书都是千篇一律的。定义、定理、公理、推论、公式，无一不是用黑体字凸显在满篇的数字字母文字的叙述中，只有这本《爱上统计学》，它甚至可以告诉你，字母不是绝对的，你喜欢用别的字母那就去用吧，名词不是绝对的，你喜欢用别的名词那就去用吧。案例都是吸引人而且切题的，讲解很简单但是很容易懂，这是一本没有被老师选中的好教材。

——当当读者741881752@***.***评

很好的一本统计学入门书

很适合统计学小白，再不必为课本上一堆发晕的数据和术语头大了。我上过概率统计的课，也看过很多做统计的论文，但是经常被那些枯燥或者“严谨”的推论吓倒，终于发现了这本书，饶有兴趣而且浅显易懂，有了这本书的基础，对深入了解别的概念很有帮助。

——卓越读者castourism评

开心的学习

对统计学感到头疼但又不得不学的我来说，这本书真是太适合了。本书的写作语言易懂，鲜有大片大片的数学公式出现，把道理都讲得很清楚。我一拿到书，边看边用软件练习，感觉非常愉快，一不小心就看了一半。真是有点“爱上统计学”了。

——当当读者leewanbo评

资源地址

上架建议：学术社科

ISBN 978-7-5624-9761-5

9 787562 497615 >

定价：58.00元

万卷方法

爱上统计学：Excel 2010 Edition

AISHANG TONGJIXUE

（美）尼尔 · J.萨尔金德 著
史玲玲 译

重庆大学出版社

Statistics for people who (think they) hate statistics: Excel 2010 edition / Neil J. Salkind. —3rd ed.

English language edition published by SAGE Publications Inc., A SAGE Publications Company of Thousand Oaks, London, New Delhi, Singapore and Washington D.C., © 2013 by SAGE Publications, Inc.

版贸渝核字(2010)第 16 号

图书在版编目(CIP)数据

爱上统计学:Excel/(美)尼尔·J.萨尔金德(Salkind, N.J.)著;史玲玲译.—重庆:重庆大学出版社,2019.8

(万卷方法)

书名原文:Statistics for People Who (Think They) Hate Statistics:Excel 2010 edition

ISBN 978-7-5624-9761-5

Ⅰ.①爱… Ⅱ.①萨… ②史… Ⅲ.①统计学—研究 ②表处理软件—应用—统计分析 Ⅳ.①C8

中国版本图书馆 CIP 数据核字(2016)第 086087 号

爱上统计学:Excel

AISHANG TONGJIXUE:Excel

(美)尼尔·J.萨尔金德(Salkind, N. J.) 著

史玲玲 译

策划编辑:雷少波 林佳木

责任编辑:陈 曦 版式设计:林佳木

责任校对:关德强 责任印制:张 策

*

重庆大学出版社出版发行

出版人:饶帮华

社址:重庆市沙坪坝区大学城西路 21 号

邮编:401331

电话:(023) 88617190 88617185(中小学)

传真:(023) 88617186 88617166

网址:http://www.cqup.com.cn

邮箱:fxk@cqup.com.cn (营销中心)

全国新华书店经销

重庆市正前方彩色印刷有限公司印刷

*

开本:787mm×1092mm 1/16 印张:19 字数:464千

2019 年 8 月第 1 版 2019 年 8 月第 1 次印刷

ISBN 978-7-5624-9761-5 定价:58.00 元

作者简介

尼尔·J.萨尔金德(Neil J. Salkind)是马里兰大学人类发展学的博士,在堪萨斯大学从教35年之后仍然是心理学和教育学研究系的荣誉教授,也一直与同事、学生保持着合作。他的早期兴趣是儿童认知发展,在完成一些认知风格和儿童多动症领域的相关研究之后,他进入北卡罗来纳大学布什儿童和家庭政策中心做博士后。他的工作发生了转向,重点转向儿童和家庭政策,特别是不同公共支持方案对儿童和家庭产生的影响。他已经发表150多篇专业的文章和报告,写了100多本普通图书和教科书,也是《爱上统计学》[*Statistics for People Who (Think They) Hate Statistics*](Sage出版社),《人类发展理论导论》(*Theories of Human Development*)(Sage出版社),《如何进行研究》(*Exploring Research*)(Prentice Hall出版社)的作者。他主编了几本专科全书,例如《人类发展全书》(*Encyclopedia of Human Development*),《测量与统计全书》(*Encyclopedia of Measurement and Statistics*)。最近出版了《研究设计全书》(*Encyclopedia of Research Design*)。他担任《儿童发展研究摘要和目录》(*Child Development Abstracts and Bibliography*)编辑13年。他住在堪萨斯州劳伦斯市,喜欢读书、与"鲨"*共泳、做布朗尼(食谱见附录),以及修理老旧的沃尔夫和老旧的房子。

---

* "鲨"指的是堪萨斯州劳伦斯市的River City Sharks游泳俱乐部。

给学生的话：我为什么写这本书

帮助学生学习新的技能，写作《爱上统计学：Excel》（重点是使用 Excel 做统计分析），对我来说都是乐趣，我也希望你们的学习过程充满乐趣，并有所收获。

学习统计学的学生和从事其他职业的人（研究者、行政人员、不同领域的专家）的共同点是相当焦虑，焦虑的原因通常来自他们从其他学生那里**听来**的经验之谈。一般来说，他们听到的一小部分是正确的——学习统计学要投入很多的时间和精力（而且偶尔还会遇到“怪物”老师）。但是他们听到的大部分（这是他们焦虑的主要原因）——统计学特别难学，令人困扰——是不正确的。实际上大多数被吓倒的学生都通过了他们以为不能通过的课程。只要集中精力，按部就班，将基本原理应用于真实的生活来理解，他们都能通过这门课程，甚至能够在这个过程中得到乐趣。这是我在写《爱上统计学》前两版时尽力去做的，在这一版的修订中我会更加努力来做到这一点。

在经过不断地摸索以及少量成功大量失败的尝试之后，我已经学会了以某种方式教授统计学，我和我的许多学生认为这种方式不会让人感到害怕，同时能够传递大量的信息。我已经尽了最大的努力将所有的经验都吸收到这本书中。

通过这本书可以了解基础统计学的范围并学习所有应该掌握的知识，也可以了解整理和分析数据的基本思路和最常用的技巧。本书有一些理论内容，但是很少，数学证明或特定数学公式的原理讨论也很少。

为什么《爱上统计学》这本书不增加更多理论内容？很简单，初学者不需要。这并不是我认为理论不重要，而是在学习的这个阶段，我想提供的是我认为通过一定程度的努力可以理解和掌握的资料，同时又不会让你感到害怕而放弃将来选修更多的课程。我和其他老师都希望你能通过这门课程。

因此，如果你想详细了解方差分析中 F 值的含义，可以从 Sage 出版社查找其他的好书（很乐意给大家推荐书目）。但是如果你想了解统计学为什么以及如何为你所用，这本书很合适。这本书能帮助你理解在专业文章中看到的资料，解释许多统计分析结果的意义，并且能教你运用基本的统计方法。

如果你想讨论教授或学习统计学的任何问题，可以随时和我联系，我的学校 E-mail 是 njs@ ku.edu。你也可以在 statisticsforpeople 网上查到《爱上统计学》所有版本的最新信息。

祝大家好运，希望你们能让我知道如何修订这本书才能更好地满足初学统计学的学生的需求。

给老师的话

这一版，我第一次在书中增加给教师的话，我主要想分享两点。

第一，感谢你们为统计学教学所做的努力。对一些学生来说这些内容比较容易，但大多数学生会认为具有挑战性。所有的学生都会感谢你的耐心和努力，这个过程中如果我可以做什么，请告诉我。

第二，《爱上统计学》并不意味着它是你们曾经看到的同类型教材的简化版。书名仅仅是传递这样的信息——许多初学统计学的学生对于即将要学的内容非常焦虑。这本书不是写给笨蛋或傻瓜的学术著作。我尽力以学生应得的尊敬来对待学生而不是施恩于他们，也尽力保证所选的学习内容是适当的。我在这两个方面做得如何，得由你们评价。但是我想清楚地表达我的意图和情感，即这本书所包含的内容是一门入门课所需要的，也许我的表述有些许幽默，但是目的本身是非常严肃的。谢谢。

致　谢

每一个人，我是说在 Sage 出版社工作的每一个人[包括投递服务部的斯蒂夫(Steve)、编辑部的卡利(Kalie)和市场部的妮科尔(Nicole)]都应该得到衷心的感谢，感谢他们的支持、建议和专业修养。这样的专业修养让他们采纳了我的一个小想法(那是在第一版出版以前)，使其变成大家现在正在阅读的这本书，并且很成功。

同时这里必须特别感谢一些人的关注和辛勤工作。德博拉·劳顿(C. Deborah Laughton)支持了写这类书的最初想法，莉萨·奎瓦斯·肖(Lisa Cuevas Shaw)鼓励撰写以 Excel 作为统计学入门教学辅助软件的图书。维基·奈特(Vicki Knight)作为出版者看到了这一版成功的前景，而且她提供了必要的支持和耐心使其成为现实。她的经验、洞见和鼓励是每个作者期望遇到的。我非常感激。其他需要特别感谢的人是副编辑劳伦·哈比卜(Lauren Habib)、编辑助理卡利·卡斯切尔拉克(Kalie Koscielak)、市场经理妮科尔·埃利奥特(Nicole Elliott)和制作编辑莉比·拉森(Libby Larson)。特别感谢宝拉·弗莱明(Paula Fleming)，她有锐利的眼光和很好的平面排版技术，使得这本书以现在这种易读的形式呈现。

同时感谢以下的人员，感谢他们对于之前版本和这一版的反馈。也向我遗漏的人员表示歉意。

Stacey Barlow Hills, Utah State University

Deborah Behan, University of Texas at Arlington

Teresa Martinelli-Lee, University of La Verne

Chris O. Odionu, Alabama A & M University

Andrew Tinsley, Eastern Kentucky University

现在,关于这一版……

以上的内容说明了我写这本书的最初想法,但是关于新的版本介绍很少。

任何一本书都需要不断修订,《爱上统计学:Excel》也不例外。过去的五年许多人曾告诉我这本书是多么的有帮助,但是其他人也告诉我他们多么想让这本书做出修订及其原因。在修订的过程中我尽力满足这两部分人的需求。书中部分内容保持不变,部分做了修订。

总是有新的事物值得关注,同时也需要用不同的方式重新理解老的主题和观念。以下的清单是《爱上统计学:Excel》中的新内容。

- 在每一章的最后增加了新的练习题。不是简单的数量增加,每一道题在应用水平和趣味(我希望大家有)方面都有很大的变化。这些练习题使用的数据集可以扫本书所附二维码下载*。本书使用的数据是用 Excel 2010 建立的,但是这些数据文件也可以在 Excel 之前的版本上应用。
- “练习时间”中问题的答案放在本书单独的部分——附录 D。经过学生和教师之间的几次争论,目前的位置似乎更好。
- 新增的第 10 章的内容是单样本检验,补充了上一个版本的缺漏。这一章对推论检验做了简明扼要的介绍。
- 另外,第 8 章讨论推论统计的开始部分增加了更多关于中心极限定理的内容。
- 本版以 Office 2010 软件中 Excel 作为重点。这个版本的 Excel 有一些重要的变化,大家可以看附录 A。附录 A 是主要变化索引。例如,许多 Excel 函数已经重新命名。其中的许多名称非常类似(例如 STDEVP 改为 STDEV.P,我们应付得来这个)。
- Mac 系统和 Windows 系统的 Excel 几乎是一样的,也非常容易互用,这本书的使用者可以使用任何一种平台,但存在一个巨大差别——Mac 版本还不能提供便宜的分析工具库(而整本书都会用到这个)。真是的,微软在想什么? 函数和公式都可以使用(但是就会少了很多乐趣)。数据文件完全可以在这两个系统之间互用。

本书中出现的文字错误或类似错误都是我的责任,我向因这些错误而带来不便的老师和学生道歉。我非常感谢所有那些指正错误并使得本版更好的信件、电话和电子邮件。在这一版,我们每个人都尽力修改之前的错误,并希望工作做得更好。期望能收到大家的建议、批评和意见。祝大家好运。

尼尔·J.萨尔金德

堪萨斯大学

njs@ ku.edu

---

* 可扫描本书封底所附二维码下载数据集以及例题中的数据。

简明目录

目 录

第Ⅰ部分 耶！我在学统计学

第Ⅱ部分 西格玛·弗洛伊德和描述统计

第Ⅲ部分 抓住那些有趣又有利的机会

第Ⅳ部分 显著性差异——使用推论统计

第Ⅴ部分 你得了解和记忆的内容

第Ⅰ部分

耶！我在学统计学

这有什么好欢呼的，你会这样说吗？现在给我几分钟的时间向你展示成功的科学家如何使用被广泛应用的叫作统计学的工具。

- 米歇尔·兰普尔（Michelle Lampl）是艾莫里大学的儿科专家和人类学家。她和朋友喝咖啡的时候，朋友谈到她的孩子长得多么快。实际上，这个初为人母的朋友几乎是说她的儿子“像野草一样疯长”。兰普尔博士十分好奇（像所有的科学家对事物感到好奇一样），她想她应该实际测量这个孩子以及其他孩子在婴儿期的生长速度。她着手测量一群孩子每天的生长，让她吃惊的是她发现一些婴儿差不多一个晚上就长一英寸！确实是爆发式生长。

想知道更多吗？请阅读原文。你可以从兰普尔、费尔德伊斯和约翰森（Lampl, M., Veldhuis, J.D., and Johnson, M.L., 1992）发表在《科学》（*Science*）258 期 801-803 页上的文章“突变和停滞：人类生长模式（Saltation and stasis: A model of human growth.）”中了解更多。

- 苏·肯珀（Sue Kemper）是堪萨斯大学心理学教授，曾经研究许多非常有趣的项目。她和其他的研究者正在研究修女群体，分析她们的早期经验、活动、人格特征和其他信息与她们中年以后的健康状况之间的关系。最特别的是这个由不同科学家组成的小组（包括心理学家、语言学家和神经学家，等等）想知道所有信息多大程度上可以预测老年痴呆症的发生。她发现修女在 20 多岁时写作的复杂性和她们在 50、60 年或者 70 年后患上老年痴呆症的可能性有关。

想知道更多吗？请阅读原文。你可以从斯诺顿、肯珀、莫蒂默、格雷纳、韦斯坦因和马克斯贝里（Snowdon, D.R., Kemper, S.J., Mortimer, J.A., Greiner, L.H., Wekstein, D.A., and Markesbery, W.R., 1996）发表在《美国医学协会杂志》（*Journal of the American Medical Association*）275 期 528-532 页上的文章“早期生活的语言能力和认知功能与晚期生活的老年痴呆症：修女研究的发现（Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the nun study.）”中了解更多。

- 阿莱莎·休斯顿（Aletha Huston）是德州大学奥斯汀分校的研究者和教师，她投入大量的研究来分析看电视对幼儿心理发展的作用。其中一个工作是她和她的前夫约翰·赖特（John.C.Wright）调查入学前观看一定量的电视教育节目对入学后学习的

影响。他们发现了确凿的证据可证明，观看教育节目如《罗杰斯先生》或《芝麻街》的儿童比没有看的儿童在学校表现得更好。

想知道更多吗？请读原著。你可以从柯林斯、赖特、安德森、休斯顿、施密特和麦凯尔伊（Collins，P.A.，Wright，J.C.，Anderson，R.，Huston，A.C.，Schmitt，K.，and McElroy，E.，1997）提交于在美国阿尔伯克基举行的儿童发展研究协会年会的论文"幼儿时期媒体教育对青少年学习成绩的影响（Effects of early childhood media use on adolescent achievement.）"中了解更多。

所有的研究者都有一个特别的问题，他们对这个问题有兴趣并用他们的直觉、好奇心和学术技能来回答这个问题。作为调查的一部分，他们使用我们称作统计学的工具分析所收集到的所有资料的意义。如果没有这些工具，所有的资料就是不相关的资料的汇集。那么这些资料就不能在兰普尔的研究中用于得出儿童生长的结论，不能在肯珀的研究中对老年痴呆症有更好的解释，也不能在休斯顿和赖特的研究中更好地理解看电视对幼儿学习和社会发展的影响。

统计学——整理和分析资料以使得资料更容易理解的科学，使得研究任务可行。通过这样的研究所得出的任何结论都是有用的，原因是我们使用统计学使得这些结果有意义。这也正是本书的目标——让你理解这些基本的工具以及这些工具的用途，当然也包括如何使用这些工具。

本书的第Ⅰ部分介绍统计学学习的内容，以及为什么值得花费精力来掌握这个领域关键的基本术语和思路。这都是为学习本书其他部分做准备。

第1章后面的两个章节我们会安排进入Excel学习，1A介绍公式和函数，1B介绍分析工具库的使用。本书使用Excel 2010（Mac爱好者是2011版）。虽然你可以使用其他版本来分析本书提供的资料，但是若使用2010这个版本，你会发现在跟紧学习进度、完成练习题方面更加得心应手。

统计学还是虐待学？这由你决定 1

本章你会学到什么

- 统计学的学习内容
- 为什么学习统计学
- 如何通过这门课程的考试

为什么学习统计学

你以前肯定听到过有人抱怨“统计学很难学”“统计中相关的数学部分很难应付”“我不知道怎么用电脑”“学统计学有什么用”“接下来不知道做什么”，还有就是统计学导论课上学生的呐喊：“我就是不懂！”

好啦，放轻松。学习统计学导论的学生总会发现他们时不时有这些想法，如果他们不和其他学生或者他们的爱人、同事或者朋友交流的话。

而且，不是开玩笑，一些统计学课程很容易被描述为虐待学。这是由于那些书无一例外的枯躁乏味，而且作者没有想象力。

而你不用忍受这些，你或者你的老师选择了《爱上统计学：Excel》，这表明你们已选择正确的统计学习方法——不会让人害怕、信息充分并且很实用（甚至有趣）的方法，而且尽可能地教你，应用统计学这一有价值的工具需要知道些什么。

如果选用这本书作为教材，这也意味着你的老师明显是站在你的立场上的——他或她知道统计学是吓人的，但他们已经采取措施来确保统计学不会让你害怕。事实上，我敢打赌你在几个星期之后喜欢上这门课程的可能性很大（这同样也难以让人相信）。

为什么使用 Excel

很简单，Excel 是目前最流行的、功能强大的应用型电子表格工具，而且在学习统计学应用的过程中它可以成为非常有用的工具。事实上，许多统计学初级课程使用 Excel 作为首要的计算工具，而不是其他的计算机软件，如 SPSS 和 MiniTab。我们不打算教你如何使用 Excel（该软件的新手可以参考附录 A），但是会展示如何使用 Excel 让你的统计学学习经历变得更好。

但和其他整理分析数据的软件一样，Excel 不是解决所有问题的灵丹妙药。Excel 有自身的局限。除非你是软件方面的专家，能够编写程序让 Excel 具有其他统计软件具有的

功能（你可能使用的程序语言是 Visual Basic Applications 或者 VBA），否则 Excel 就不像其他统计分析的软件一样完备，不能提供同样多的选项。但是在统计学初级分析的水平上，Excel 是非常强大的工具，可以完成很多统计分析功能。

开始新的内容之前我们需要了解 Excel 的一些术语。第一个类似 Excel 的计算机软件是 Visicalc[感谢丹·布利克林（Dan Bricklin）和鲍勃·弗兰克斯顿（Bob Frankston）]，以电子数据表闻名。当然，Excel 的应用也以电子数据表闻名，但是每个活动表称作工作表（worksheet）。如果多个工作表组合起来就构成了工作簿（workbook）。很有趣吧！

5 分钟统计学简史

在阅读更多内容之前，有必要从历史的视角了解统计学。你知道，毕竟几乎所有的社会科学、行为科学、生物学的大学生和教育学、护理学、心理学、社会福利和社会服务以及人类学的研究生都需要选修这门课。了解"统计学应用领域从何处开始"是否有必要？回答是"当然有必要"。

只要回顾过去，人们就会发现计数是个非常好的创意（就如"你需要多少这个来交换那个"），同样收集信息也成为有用的技能。如果用到计数，人们就会知道太阳在一季升起多少次，度过整个冬天需要多少食物，以及一个人拥有多少资源。

这只是开始。一旦数字成为语言的组成部分，似乎接下来的步骤就是将数字和结果结合在一起。在 17 世纪早期人们就开始收集与人口相关的数据。以此为基础，科学家（大多数是数学家，此外是物理学家和生物学家）需要发展特定的工具来回答特定的问题。例如，弗朗西斯·高尔顿（Francis Galton）（顺便说一下，他是达尔文的表兄弟，生于 1822 年，死于 1911 年），他对人类智力的性质非常感兴趣。为回答家庭成员智力一致性的基本问题，他使用了特定的统计工具——相关系数（最早是数学家开发的），接着他把相关系数广泛地应用于行为科学和社会科学。

你会在第 5 章对这个工具有全面的了解。实际上，你将学到的大多数基本统计过程最初应用于农业、天文学甚至政治学领域。在人类行为领域的应用则比较晚。

在过去的一百年中，人们在发明新方法应用旧观念方面取得了极大的进步。最简单的用于比较两个群体的均值差异的检验方法在 20 世纪初取得首次进展。在此基础上建立的技术十年之后才提出，而且得到了极大的完善。随着个人电脑和类似 Excel 的软件的应用，任何想研究这些有趣问题的人都可以使用这些先进的统计技术。

影响力深远的个人电脑的应用有利有弊。有利的方面是大多数统计分析不再需要巨大、昂贵的中央处理机。而价格不超过 500 美元的个人电脑就能满足 95%的人 95%的需求。另一方面，很少有大学生（如已经通过这门课的同学）会使用他们已有的旧数据，并且认为应用复杂的软件分析这些数据就可以得到可靠、可信和有意义的结果——这是不正确的。你的老师会说"垃圾进，垃圾出"——如果你开始就没有使用可靠、可信的数据，那么分析这样的数据所得到的结果就既不可信也不可靠。

现在，不同领域——从刑事犯罪学、地球物理学到心理学——的统计分析人员发现他们基本上使用相同的技术来回答不同的问题。当然在资料收集方面有重要的差别，但是大体上来说，根据数据（大量的数据）收集所进行的分析（多个分析），即使有所谓的差别也是非常细微的。基本上，这门课会给你提供一定的工具来理解统计学如何应用于几乎所有的学科，帮你轻松搞定那些 3~4 个学分的课程。

如果你想更多地了解统计学历史，而且想按历史发展顺序来了解，圣安赛姆学院或加利福尼亚大学洛杉矶分校的网站是个好地方。

好了，5 分钟结束，你已经了解了你需要了解的统计学的历史。下面让我们看看统计学是什么（或不是什么）。

统计学：是什么（或不是什么）

这是一本关于基础统计学，以及在不同的情况下如何应用统计学分析和理解数据的书。

就一般意义而言，统计学是描述一系列可用于描述、整理和解释资料或数据的统计工具和技术。这些数据可能是特定课程的学生的考试得分、解决问题的速度、吃某种药感到不适的病人人数、世界系列大赛每一局出错的次数或者圣达菲高级餐厅晚餐的平均价格。

在这些案例或者我们可以想到的更多案例中，都需要收集、整理、汇总和解释数据。在本书的描述性统计部分你可以学会收集、整理和汇总数据。而在了解了推论性统计的用途之后就可以学会解释数据。

什么是描述统计

描述统计（descriptive statistics）常用于整理、描述所收集数据的特征。所收集的资料有时也称作**数据集**（data set）或者就叫**数据**（data）。

例如，下面所列出的就是 22 名大学生的姓名、主修专业和年龄。如果需要描述大学最流行的专业是什么，你可以使用描述统计来概括他们的选择（也称作众数）。在这个例子中最大众化的专业是心理学。如果你想知道学生的平均年龄，可以很容易地计算另一个描述统计值（也叫作均值）来确定这个变量。这两个简单的描述统计值常用于描述数据。就如我们对下面 22 个学生情况的分析一样，描述统计使我们能很好地描述大的数据集的特征。

姓　名	专　业	年　龄	姓　名	专　业	年　龄
Richard	教育学	19	Elizabeth	英语	21
Sara	心理学	18	Bill	心理学	22
Andrea	教育学	19	Hadley	心理学	23
Steven	心理学	21	Buffy	教育学	21
Jordan	教育学	20	Chip	教育学	19
Pam	教育学	24	Homer	心理学	18
Michael	心理学	21	Margaret	英语	22
Liz	心理学	19	Courtney	心理学	24
Nicole	化学	19	Leonard	心理学	21
Mike	护理学	20	Jeffrey	化学	18
Kent	历史	18	Emily	西班牙语	19

你瞧，这是多么简单！确定最常被选择的专业，就是确定哪一个专业出现的次数最多。要确定平均年龄，就是将所有的年龄值加起来然后除以 22。你是对的——出现次数最多的专业是心理学（9 次），平均年龄是 20.3。瞧，不用人帮忙——你就是统计分析师。

什么是推论统计

推论统计(inferential statistics)通常是(但并非总是)数据收集和汇总后的下一步。推论统计常利用较小群体的数据(如我们的22个学生构成的群体)来推论可能的较大群体(如艺术和科学院的所有学生)的特征。

这个较小的群体通常叫作**样本**(sample),是**总体**(population)的一部分或一个子集。例如,新泽西州纽瓦克的所有五年级学生构成一个总体(所有正好是五年级而且住在纽瓦克的人,都包括在内),从中选取150人就构成一个样本。

咱们来看另外一个例子。作为新雇用的研究人员,你的市场代理要你确定以下名称中的哪一个最适合用作薯片的新品牌,是薯片王、乐趣薯片,还是嚼嚼乐?作为专业统计分析人员(我知道我们现在超前了一点,但是请保持信念),你需要选取一个薯片食用者小群体,这个群体可以代表所有喜欢吃薯片的人,接着询问这个群体最喜欢这三个名称中的哪一个。如果你做得好,就可以很容易地将这个发现推论到更大的薯片食用者群体中。

或者,如果说你对某种疾病的最佳治疗方案感兴趣。你可以让一组人试用一种新药,另一组食用安慰剂(众所周知的没有任何效果的药剂),第三组什么药也不吃,然后来看结果。好吧,你会发现大多数病人在没有接受任何治疗情况下会好转,机体自身会恢复!药物没有任何效果。那么依据你的实验结果,你可以将这些信息推论到更大的遭受这种疾病痛苦的群体。

换句话说……

统计学是帮助我们理解周围世界的工具。这是通过整理我们收集到的数据实现的,而且接着还可以让我们做出特定的推断,也就是怎样将那些数据的特征应用到新的情境。描述统计和推论统计可以一起发挥作用,使用哪一种、何时使用取决于你想要解答的问题。

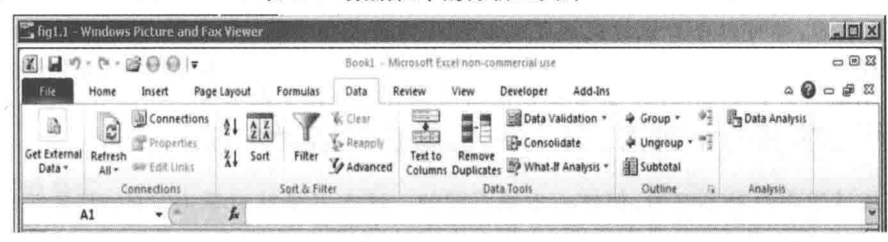
现在,掌握统计学知识比以前更重要,因为统计学帮助我们根据经验(或者观察)资料来做决策,而不是根据偏见或者信仰。比如你想了解早期干预项目的效果,那么检验这些项目是否发挥作用,并提供证据给委员会,这样他们才能够根据项目的有效性依照规定支付相关费用。

分析工具库中的工具

关于Excel你需要了解的很多基本内容可以见附录A。但是某些特定的Excel程序只有安装了分析工具库(Analysis Toolpak)(本书多个章节中会使用到)才可以使用。

分析工具库是一个非常好的Excel功能设置,它是一个特别的工具集,最初安装的Excel可能不会显示。如何知道你的Excel是否加载了这一项?如果在Excel界面的[数据](Data)的菜单中没有[数据分析](Data Analysis)显示,见表1.1,就意味着你需要安装这个加载项。

表 1.1 数据栏中的分析工具库



可以按照下面的步骤安装。

1.单击[文件](File)菜单,接着单击[选项](Options);

2.单击[加载项](Add-Ins),接着在加载项的选项框中选择[分析工具库](Analysis ToolPak);

3.单击[转到](Go);

4.在可用[加载宏](Add-Ins)选项框,单击[分析工具库](Analysis ToolPak)选项框,接着单击[确定](OK)。

之后,分析工具库可以保持一直出现在数据菜单中。你现在已经安装完毕,你的 Excel 的运用将更加富有成效并充满乐趣。你可以在 1B 小节学习如何使用分析工具库。

我在统计学课堂上做什么

你阅读这本书可能有多种原因。也许是选修了统计学导论课程,或者是在申请综合考试,甚至是在暑假预习新课(学霸!)并且想申请更高级的课程。

总之你都是学习统计学的学生,不论你是否参加课程结束后的考试,或者只是出于自己的兴趣选择这门课程。但是学习这门课有许多好的理由——它很有趣,它很重要,并且二者兼具。下面是我的学生在我们的统计学导论课程开始时听到的一些说法。

1.“统计学 101”“统计学 1”或者是你们学校用的任何其他课程名称,在你的成绩单中看起来都很重要。说真的,统计学是你完成学业的必修课程。即使不是,拥有统计学技能绝对是找工作或进一步深造的重要附加值。而且如果选修了更高级的统计学课程,你的简历肯定会更令人印象深刻。

2.如果不是必修课程,选修基础统计学可以把你和没有选修的同学区别开来。这表明你愿意选择挑战平均难度以上的课程。随着政治和经济世界变得更加“需要有说服力”,分析技能会得到更多的关注。谁知道呢,也许这门课就是你找到工作的“关键一票”。

3.基础统计学可能是你不熟悉的一种智力挑战。这个过程需要你运用一些数学计算以及理论与实践的结合。最起码,把所有活动结合起来看就是一次充满活力的智力体验,因为你要学习的是全新的领域和学科。

4.毫无疑问,如果具备一定的统计学背景,你会成为社会科学或行为科学中更优秀的学生,因为你可以更好地理解杂志中的文章,也可以更充分地理解老师和同伴在课堂内外讨论的内容或所做的研究。你可能会惊讶地第一次对自己说:“哇,我真的听懂了他们在讨论什么。”而且这会经常发生,因为你已经具备了正确地理解科学家如何得出结论所必备的基本技能。

5.如果你计划获得教育学、人类学、经济学、护理学、社会学或其他社会科学、行为科学

和生物科学领域的任何一个学科的硕士学位，统计学课程都是你前进的基础。

6.最后，你可以夸口说你完成了人人都认为是相当于建立和运行核反应堆那么高难度的课程。

使用这本书的十种方式(同时也是在学统计学啦!)

统计学书总是不好看，但这一本是不同的。这本书是针对学生写的，但是并没有降低标准，而是内容充实，也尽可能夯实基础。本书没有假定在课程开始之前应该具备什么知识，只是进程较缓、步骤较小，可以让学生按自己的节奏安排学习。

大家都认为统计学是很难掌握的课程。我们也这么认为，因为统计学中有的部分的确充满挑战。然而，无数的学生已经掌握了这门课程，那么你也可以。在开始我们的第一个主题之前，先看一下导论这一章的要点。

1.**你不笨**。这是真的。如果你是笨蛋，你不可能在学业方面走这么远。因此，对待统计学就像对待其他新课程那样吧。听讲座、学习基本内容、做书上的练习或课堂练习，那样你就会学得很好。火箭研究者精通统计学，不过你不需要像火箭研究者那样研究统计学。

2.**你怎么知道统计学很难学?** 统计学很难学？既是也不是。如果你是从上过这门课的朋友那儿听来的，而他们没有努力学习也没有学好，那他们肯定会告诉你统计学是多么难学，甚至会说，统计学即使不是整个生活的灾难，也是整个学期的灾难。不过不要忘记——抱怨多来自失意者。因此，我们建议你应该以这种态度开始这门课程，也就是：等着看统计学是否难学，并且依据自己的经验做出判断。最好是找几个上过这门课的人讨论一下，获得他们的总体看法。不要仅依据失意者的经验做出判断。

3.**不要逃课——按顺序学习各章的内容**。本书的每一章都是下一章的基础。我希望在课堂上学习了所有内容后，你就可以回顾整本书，并把这本书作为参考书。如果你需要一个表中的特定值，可以查阅附录B。或者，你需要记起如何计算标准差，可以回顾第3章的内容。但是现在要按照本书的顺序学习每一章。当然也可以不按照顺序，先了解后面的学习内容。但是在学习后面章节之前务必掌握前面的章节。

4.**形成学习小组**。这是确保通过这门课程的最基本的方式之一。在一个学期的开始，要和朋友约定一起学习。如果没有朋友选择同一课程，那么就结交新朋友，或者邀请和你一样看起来很乐意学习统计学的学生一起学习。如果你学得比他人好，一起学习可以帮助他人；反之，你就可以从他人那里受益。每个星期安排特定的时间聚在一起一个小时，复习每一章后的练习题，或者相互提问。也可以依据需要安排更多时间。与他人一起学习是帮助你理解和掌握课程内容的有效方式。

5.**向老师提问或者向朋友提问**。如果你不了解课堂上所讲授的内容，可以向老师提问来弄清楚。毫无疑问——如果你不理解课程内容，那么这很可能是很多人共同的问题。老师一般都欢迎提问。特别是如果你上课前已经预习，你的问题就包含更多信息，可以帮助其他学生更好地理解课程内容。

6.**完成章后练习题**。练习题是以每一章的内容和案例为基础的。这些练习题可以帮助你运用每一章讲授的概念，同时可以建立自信心。这些练习题如何发挥作用？每一个练习题的解答都对应着一个统计学要点，答案见附录D。如果可以解答章后的练习题，那么你已经很好地掌握了这一章的内容。

7.练习，练习，再练习。是的，这是个很古老的笑话。

问：如何才能到卡内基大厅*？

答：练习，练习，再练习。

好吧，这和基础统计学没有区别。你必须应用学到的知识，而且通过经常应用来掌握不同的方法和技术。这意味着解答1~16章的练习题，好好利用遇到的任何机会帮助自己理解学到的内容。

8.寻找应用实例使得所学内容更现实。在其他课上，你可能有机会读到杂志上发表的文章，讨论研究的结论，并且对你的学习领域的科学研究方法的重要性进行一般的讨论。这些都是机会，可以考察所学的统计学知识如何帮助你更好地理解课堂讨论的主题和应用统计学的领域。这个想法实践得越多，你的理解就越好、越充分。

9.浏览。首先阅读指定的章节，然后返回来带着目的阅读。选择舒服的方式来考察《爱上统计学：Excel》每一章包含的内容。不要逼迫自己。了解将要学习的主题的同时熟悉现在的课堂内容总会有好处。

10.乐在其中。这听起来好像很奇怪。但是总的来说，与其让这门课和它的要求控制你，不如你控制这门课。建立学习计划并按计划完成，在课堂上提问题，而且将智力实践看作是成长的一种形式。掌握新知识总是令人激动和满足——这是人类活力的体现。在这里你也可以体验这种满足感——集中精力、做出达到某个成绩水平的承诺并且努力学习。

Mac 使用者注意事项。微软设计人员这些年来对 Mac 平台变得更为友好。Windows 系统和 Mac 系统的 Excel 最新版操作系统几乎是一样的。较大的差异（实际上不是很大）是完成特定任务时所使用的按键。例如，Windows 系统使用 Ctrl+C 复制选中的文件，Mac 系统使用 Apple 键或者 Command 键（键盘左下侧有四个环形曲线标志的很小的键）加 C 键完成同样的复制功能。信不信由你，Apple 键也被叫作长条键（splat）、四叶草键（cloverleaf）、蝴蝶键（butterfly）、小帽键（beanie）和花键（flower）。不论在哪个操作系统中使用 Excel 都差不多，在适应方面不会遇到困难。而且 Windows 版或者 Mac 版的 Excel 都能互读文件，可以很安全地在两个操作系统之间转换文件。如果实在想要完全一致，Mac 用户可以进入 System Preferences 重新设置键盘，确保 Windows 和 Mac 的按键完全一样。

两个版本之间最大的差异是 Mac 版不可以使用工具库（ToolPak）。我们期望以后可以使用，但是 2011 版还不可以。

关于那些符号

一个符号是一种象征。浏览全书你会发现有许多不同的符号。下面是每一个符号及其所代表的意义。



这个符号表示正文之外的信息。我们发现有必要详细说明特定的观点，而且我们能够很容易地在常规内容之外做到这一点。



我们可以选择一些更专业的观点，然后简要地讨论并告诉你哪些内容超过课程的范围。你会发现这很有意思，也很有用。

---

* 美国有名的音乐演奏厅，只有达到相当高演奏水平的音乐人士才有资格在那里登台。这里的问句本来是问路。——译者注


浏览《爱上统计学：Excel》全书你会发现许多小梯子符号，就像在这儿看到的。这表明这里有许多步骤，指导你进行特定的过程。这些步骤都已经通过相关机构的验证。


绑着蝴蝶结的手指是很可爱的符号，但它的主要目的是用于强调你所读到主题的重点。在学习过程中，强调某些点是由于这些重点通常是这个主题的关键。


《爱上统计学：Excel》的大多数章节提供了一个或多个特殊的统计过程和同步计算的详细信息。计算机符号用于表明每章“使用计算机……”的部分。

更多 Excel 符号表明这个内容是已经提过或者正在应用的 Excel 知识。

附录 A 是额外的 Excel 学习，包括你应该了解的 Excel 的知识，以及人人都应该掌握的 50 个重要的 Excel 基本技巧。

附录 B 包含你会学到的重要的表格，在全书都会用到。

附录 C 的数据集用来完成本书的练习题。你会发现数据集以类似“第 2 章数据集 1”这样的方式命名，每一个这样的数据集都在附录 C。你可以从出版社提供的网址直接下载，也可以向邮箱 wjffsyh@ foxmail.com 发邮件索要。

难度指数

为了帮助你跟进学习进度，我们在每一章的开头设置了难度指数。这为每一章的开头增加了乐趣，而且也是非常有用的提示，让你知道将要面对什么，与其他章节相比这些章节的内容难度如何。

1 非常难 ☺
2 比较难 ☺☺
3 一般 ☺☺☺
4 比较容易 ☺☺☺☺
5 非常容易 ☺☺☺☺☺

“多少 Excel”指数

为了帮助你更好地跟进学习进度，我们在每一章的开始还设置了“多少 Excel”指数。这为每一章的开始增加了更多乐趣（或者痛苦），但是也让你知道在这一章将要面对多少 Excel 内容。

（很少）
（一些）
（很多）
（非常多）
（超级多）

术　语

正文中一些重要的术语都汇总在本书最后的术语表中，并进行了简单的解释。

小 结

实际情况不是那么糟，对吧？我们想要鼓励大家继续学习，并且不要担心难易程度和时间花费，以及能否自如应用。每次只学习一章，就像学习这章一样。

练 习

为了帮助你回顾每一章覆盖的内容，1～17 章的末尾都提供了习题集，本书附录 D 提供了练习题的答案。举例示范一下，下面就是第一个习题集（但是不要企图看答案，因为答案是“自己解决”那类回答——每个答案都与你自己的经验和兴趣高度相关。）

1. 访问一个日常工作中应用统计学的人。这个人可能是你的顾问、指导教师、生活在同一街区的研究者、公司的市场分析人员或城市规划人员。询问他或她最喜欢的统计课程。了解他或她喜欢什么不喜欢什么。他或她是否给你一些有助于成功的建议。而且最重要的是，询问他或她如何在工作中使用这一对你而言全新的工具（统计学）。
2. 我们希望你是某个学习小组成员，但如果不是，你可以寻找电话、邮件、QQ 或者微信的学习伙伴（甚至不止一个）。与你的学习小组或学习伙伴讨论类似讨厌或者害怕统计学课程等问题。你们的共同点是什么？不同点是什么？与你的学习伙伴讨论克服畏难心理的策略。
3. 查阅本地报纸，找出任何主题的调查和访谈的结论。总结这些结论，然后以你的能力尽可能地描述相关的研究人员或者调查人员如何得出他们的结论。这个过程可能容易也可能不容易。一旦你了解了他们这样做的原因，尽力推测其他的收集、整理和汇总同样数据的方法。
4. 去图书馆复印一些你的专业领域的文章。然后阅读文章，重点关注整理和分析数据的统计过程部分（通常是文章的“结果”部分）。你可能对具体的统计过程了解不多，但是你能识别多少不同的统计过程（如 t 检验、均值和标准差的计算）？你是否能完成下一步的工作？并且告诉你的老师这些结果和研究问题的关系，或者与最初的研究课题的关系。
5. 在网上查找五个任何主题的数据库，然后简要介绍其所提供的数据类型以及数据是如何整理的。例如，登录网络数据的一个源头——美国人口统计局官方网站，你会发现“Access Tools”链接，进入后的网页提供真实数据的链接。尽力查找适合你的专业的数据和资料。
6. 完成这项作业有很多附加分，找到一个实际使用 Excel 进行日常工作中数据分析的人，问他或者她为什么使用 Excel 而不是其他更受重视的软件，如 SPSS 或者 MiniTab。从政治科学到护理学多个领域都非常容易发现这样的人，放手去寻找吧。
7. 最后，既然你已经在坚持完成第一个习题集，那么，提出你的专业领域或者兴趣方面的 5 个最有意思的问题。尽力提出这样的问题，因为你可能想要真实的数据和资料来回答这些问题。来做一回科学家吧！

1A 你应该了解的公式和函数

本章你会学到什么

- 公式和函数之间的差异
- 如何建立和使用公式
- 重要的 Excel 函数
- 如何选择和使用函数

Excel 神奇的工具库中最有价值的就是公式和函数。公式和函数可以让你跳过让人厌烦的计算并得到最终的结果。公式和函数都是捷径,并且工作方式不同,工作内容不同。我们先从公式开始。

什么是公式?

你可能已经知道了这个问题的答案。一个公式就是一个运算法则集,完成特定的数学运算任务。例如,下面是一个简单的公式。

$$2+2=$$

"+"号将特定的值(一个 2 与另一个 2)加起来得到结果(4)。这个公式很简单。

下面的公式有些超过了你的理解,在本书第 16 章中会有更多的解释。

$$Y' = bX + a \quad (1a.1)$$

这个公式常用于依据 b 值、X 值和 a 值来预测 Y' 值。我们之后会讨论这些符号的含义,这个公式包含了很多运算符号和数学运算法则,并可以帮助我们计算需要确定的数字。

Excel 是一个公式生产机器,方便你使用这些工具来让你的统计学学习变得更容易。



通过以下步骤建立公式:

1.单击一个单元格,设定公式结果在这个单元格中出现。

2.输入等号,也就是"="。不论公式中的内容是什么,所有的公式都以等号开始。

3.输入公式。在公式中没有空格——Excel 不识别空格。

4.按回车键,好啦!公式的结果就出现在你选择的单元格中。

例如，让我们输入之前列举的公式——2+2，看这些步骤是否有效。

1.如表1A.1所示，我们选择单元格A1。

2.输入等号，如图1A.2所示。在表头的公式栏就处于激活状态。我们输入单元格A1的所有内容都会在公式栏出现。

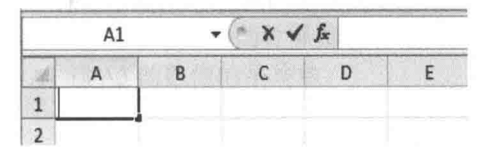


图1A.1　选择输入公式的单元格

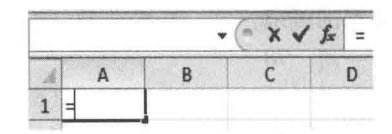


图1A.2　输入等号表示建立公式的开始

3.输入公式的其他部分，在这个例子中是(2+4)/2，如图1A.3所示。

4.按回车键，公式的计算结果就出现在A1中，如图1A.4所示。

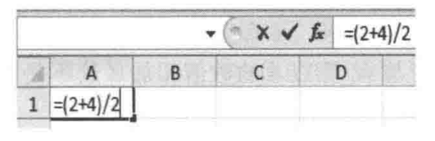


图1A.3　在单元格A1中输入公式

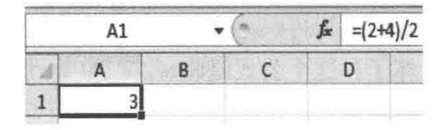


图1A.4　公式的结果就出现在A1

一些注意事项

- 公式总是以等号开始，这是告知Excel接下来输入的是公式。
- 公式本身一般是在公式栏显示。
- 公式的计算结果（不是公式本身）会回到选定的单元格中。

这是如何使用公式的最简单的案例。公式可以依据需要确定复杂程度。

更多Excel

如何在工作表中看到隐藏在背后的公式？使用Ctrl+`键组合可以实现公式和这些公式的结果之间的转换。`键是键盘上方1键左侧的键。

运算符，运算符——给我一个公式！

你已经看到，即使最简单的公式也包含多个运算法则。这个案例中的运算法则是加法和除法，这两个运算法则命令Excel将你在图1A.3中所见的数值加起来，除以2，计算结果在单元格A1中出现。

加法和除法只是两类运算法则，大多数重要的运算法则及其符号请见下表。

运算法则	运算符号	例子	公式含义
加法	+（加号）	= 2+5	2加5
减法	-（减号）	= 5-3	5减3
除法	/（除号）	= 10/5	10除以5
乘法	*（乘号）	= 2*5	2乘以5
平方	^（平方号）	= 4^2	4的2次方，或4的平方

注意括号的应用

如果你建立的公式比较复杂，考虑运算过程的次序和括号的使用就非常关键。看现在的例子，我们想计算一个月每周五的周测成绩的平均值，成绩的数值范围是 0~100。下面是威利的成绩：

第一周　78

第二周　56

第三周　85

第四周　92(威利终于取得了好成绩!)

我们需要建立一个公式将所有的数值加起来然后除以 4，我们将每个数值命名为 w_1、w_2、w_3 和 w_4。下面是我们可能建立的公式之一：

$w_1+w_2+w_3+w_4/4$

坏了！这个公式是将 w_1、w_2和 w_3 加起来，接着与 w_4 除以 4 的数值相加。这不是我们想要的结果。

我们实际上建立的公式应该如下所示：

$(w_1+w_2+w_3+w_4)/4$

就是这个公式。我们将四个数加起来然后除以 4，这个公式有效。应该总结什么经验？如果你应用的公式比较复杂就应考虑括号的应用。

什么是函数

你已经知道公式是简单的符号集(如数字和运算法则)，可以用于计算，并且将得出的结果显示在公式定义的单元格中。

函数无非就是预先进行定义的公式。设计 Excel 的人员已经建立了大量可以进行各种计算的函数，但是就我们本书的目的来说，我们只需要应用与本书章节相关的函数。这些函数都出现在功能区的[公式](Formulas)菜单下，如图 1A.5 所示。

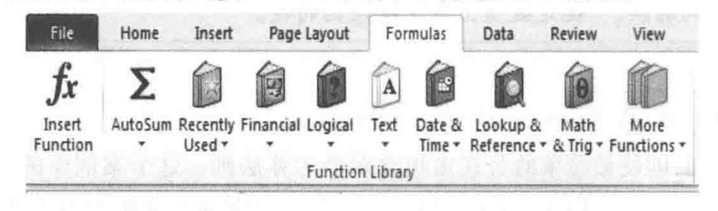


图 1A.5　Excel 功能区的 Formulas 菜单

例如，这里有财务函数、逻辑函数、文本函数和其他函数。但是我们关注的重点只是统计函数分类中的函数。只有单击[其他函数](More Functions)的下拉选项，见图 1A.5，再单击[统计](statistical)，才可以在屏幕显示统计函数。而且在所有的统计函数中我们只使用和本书内容相关的函数，例如 AVERAGE(猜猜是什么函数)、T.TEST(你可能猜不出来)。一些函数可能现在对你比较难，如 FISHER 和 GAMALIN，我们先不解释。这样的函数就留给你们到其他课程学习或者自己钻研。

使用函数

与公式不一样，函数不需要你建立。你只要输入你需要处理（在哪些单元格）的数值。每一个公式包含两个不同的要素——函数的名称和函数的参数。

下面是非常简单的计算给定数据集平均数的函数。在案例中，这个函数计算从单元格 A1 到 A3 的数值的平均数。

=AVERAGE(A1:A3)

函数的名称是 AVERAGE（平均数），函数参数是 A1:A3——这是你想用函数发挥其巧妙作用的单元格。就如你所见，函数（与公式类似）始终、一直、从头到尾都是以等号开始。

下面是另一个函数，用来计算一定数量单元格的和。

=SUM(A1:A3)

你可能会认为这很简单。为什么不在这种情况下使用公式？你当然可以使用。但是如果需要计算 3 267 个数值的和，就像下面的函数所示，那你怎么办？

=SUM(A1:A3267)

或者如果你需要进行很花哨的计算，包含非常复杂的函数怎么办？那是另外的情况，并且你真的不愿意手动输入=(A1+A2+A3+A4...)直到 A3267，对吧？我们认为不会。

现在我们学习使用函数的方式，以 AVERAGE 函数为例。

使用这个函数（或其他函数），遵循以下三步。

1.在你选定的结果出现的单元格中输入函数。

2.输入你希望函数计算的单元格的范围。

3.按回车键，好啦，结果就出现了。

但是完成这三步的方式有几种，我们现在分别来看。

插入函数（如果你知道函数的名称和工作方式）

这是老方法。

1.在你选定的结果出现的单元格中输入函数。

	A	B
1		Value
2		3
3		4
4		2
5		3
6		4
7		5
8		4
9		3
10		2
11		3
12	Average	

图 1A.6　建立数据集并确定应用 AVERAGE 函数的位置

B12　fx =AVERAGE(B2:B11)

	A	B	C	D	E
1		Value			
2		3			
3		4			
4		2			
5		3			
6		4			
7		5			
8		4			
9		3			
10		2			
11		3			
12	Average	3.3			

图 1A.7　完成 AVERAGE 函数

例如，在图 1A.6 所示，你可以看到 10 个数构成的数据集。我们使用 AVERAGE 函数计算这些数的平均数。为了让这个过程更清晰，我们在希望结果出现的单元格的左侧单

元格中录入了文本标签。

2.在单元格 B12 输入=AVERAGE(B2:B11)。

3.按回车键，结果立即出现：就如在图 1A.7 所见，结果出现在单元格 B12，而且在公式栏可以看到公式的构成。

要注意函数的结果（3.3）回到了相同的单元格（B12），也就是函数输入的单元格。很酷吧！

一点也不难，而且非常方便。记住你已经可以使用函数了。但是你如何知道函数的结构呢？这是下一步要学习的内容。

> **更多 Excel**
>
> 那么你如何知道用什么公式？好吧，稳妥的方式是通过研究不同的函数，了解函数的不同功能（通过本书了解函数）。另一种方式是使用 Excel 的 HELP 功能（在任何时候按 F1 键，并输入你需要帮助的主题词）。还有一种方式是查看这一章最后的表 1A.1，表中简要列出了本书会提到的函数以及函数的功能。

使用[插入函数](Insert Function)(fx)命令插入函数

我们使用相同的例子，就是 AVERAGE 函数，并且假定我们之前没有使用过这个函数，但是知道这是你想使用的函数。

我们使用图 1A.6 所示的相同数据。首先，删除单元格 B12 中函数的结构。删除的方式是选中这个单元格，单击空格键，接着按回车键。

1.选择单元格 B12。

2.单击[公式](Formulas)菜单和[插入函数](Insert Function)(fx)命令。你这样做之后就会看到如图 1A.8 所示的[插入函数](Insert Function)对话框。

> **更多 Excel**
>
> [插入函数](fx)命令的捷径就在公式栏的左侧，在公式栏你可以看到单元格中的所有信息。只要单击一下，你就可以得到[插入函数](Insert Function)对话框。

3.现在你可以进行下面的任何一项。

a.输入你要进行的计算的简短描述，例如 AVERAGE，然后单击[转到](Go)选项；或者

b.在函数列表中找到你需要的函数，然后双击这个函数。

我们选择方式 b，当我们这样做时，函数参数对话框就出现了，如图 1A.9 所示。注意 Excel 自动假定我们想要计算选定的单元格之上的所有数的平均数，而且在 Number1 对话框中就选定了单元格的范围。

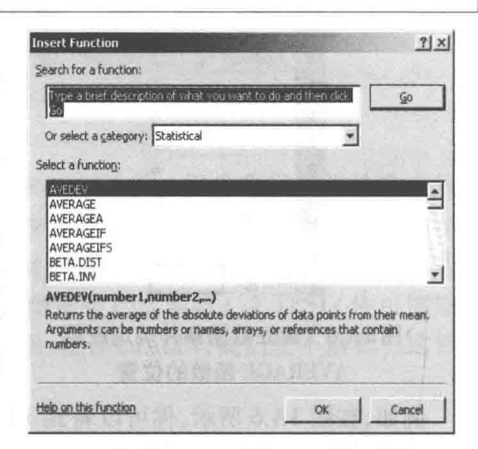


图 1A.8　插入函数对话框

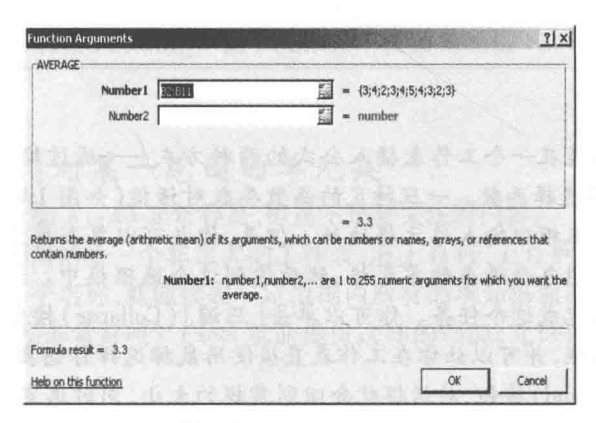


图 1A.9 函数参数对话框



别激动。函数 argument 不是真的因意见不同而引发的争论。数学用语中 argument 是一系列参数,也就是任何函数的括号中你需要提供的内容——一系列函数运行的参数。

现在我们来了解这个对话框中的不同要素。

- 函数的名称是 AVERAGE 函数。
- 在文本框中输入单元格的范围(参数),单元格的范围是你希望使用函数进行运算的数的范围。注意 Excel 非常智能,可以自动输入它认为你想计算平均数的单元格的范围。同时也注意你实际需要计算平均值的数在文本框的右侧呈现。
- 在这个对话框的中间(靠左侧)是函数功能的描述,右侧是函数运算结果,这个结果将会出现在你确定的结果所在的单元格中,在这个案例中这个值是 3.3。
- 如何将函数组合在一起的语法(命令句)在靠近底部的位置给出。
- 公式的计算结果在底部左侧再一次呈现。
- 单击[确定],你就会看到如图 1A.7 所示的相同的结果。我们使用[插入函数]命令输入函数而不是输入函数的名称,但是我们会得到相同的结果。

使用[公式]→[插入函数]→[其他函数]→[统计]

其实已经说过,只要按顺序单击三次鼠标,你就可以看到所有可用的统计函数的清单。选择任何一个函数(如 AVERAGE 函数)都会出现如图 1A.9 所示函数对话框。

更多 Excel

许多函数可以完成的功能远比显示出来的多。例如,最简单的加和函数,你输入不同的参数会得到相应的结果。Excel 函数灵活变通,非常好用。

如果输入这样的公式	Excel 会这样运行
=SUM(3,4)	数值加和等于 7。
=SUM(A2:A4)	将单元格 A2 到单元格 A4 的数值加和。
=SUM(A2:A4,6)	将单元格 A2 到单元格 A4 的数值加和再加上 6 得出总和。
=SUM(A6:A8,4)	将单元格 A6 与单元格 A8 的数值相加之和再加上 4 得出总和。

更多 Excel

现在你已经知道在一个工作表插入公式的两种方式——通过输入函数名称或者在[插入函数]对话框选择函数。一旦特定的函数参数对话框(如图 1A.9 所示)打开了,你就可以在合适的文本框中输入单元格地址。但是,你也可以单击单元格地址选项框,接着拖动鼠标覆盖你想包含进来的单元格,然后拖到这个选项框中。这很好,但是还有另外一种简洁的方式完成这个任务。你可以单击[压缩](Collapse)按钮(看起来类似▣),它会压缩整个对话框,并可以让你在工作表直接使用鼠标选择你想要使用的单元格。接着单击[扩展](Expand)按钮,对话框就会回到常规的大小,同时选定的单元格已经包括进来了。

更多 Excel

如果你在单元格中直接输入一个函数,Excel 2010 提供类似的函数清单,以及如何使用函数的提示说明,就如在图 1A.10 所示。这里我们输入=aver,当我们输入时,Excel 提供了不同的平均数函数清单,以及 AVERAGE 函数用于做什么的说明。

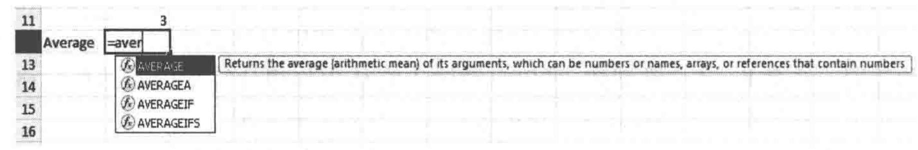


图 1A.10

在公式中使用函数

现在可以放松一下。公式和函数基本上是一回事儿,都执行我们的命令。没有理由不能将函数包含在一个公式之中。

F2　=AVERAGE(B2:D2)*E2

	A	B	C	D	E	F
1	Name	Eval 1	Eval 2	Eval 3	Fudge Factor	Final Score
2	HY	67	76	34	1.10	64.90
3	NM	65	78	32	1.02	59.50
4	GG	45	98	34	1.23	72.57
5	DF	65	98	54	1.06	76.67
6	KH	76	78	58	1.03	72.79
7	RR	32	76	75	1.04	63.44
8	YR	45	77	54	1.05	61.60
9	HH	43	76	54	1.00	57.67
10	JU	34	54	33	1.01	40.74
11	WE	32	47	78	1.04	54.43

图 1A.11　在公式中使用函数

例如,我们现在假定你有三个工作的评估得分(Eval 1、Eval 2 和 Eval 3),见图 1A.11。你也有一个经验系数(在 E 列),你可以基于你的判断用这个变量的值增加或者减少一个雇员的得分。例如,对于雇员 KH,你想将他的得分增加 3%,因此你将平均评估得分(Eval 1、Eval 2 和 Eval 3)乘以 1.03。使用 AVERAGE 函数(你会在第 2 章学习如何使用这个函数)的公式如图 1A.11 所示。

就如图 1A.11 的公式栏所示,单元格 F2 的公式如下:

=AVERAGE(B2:D2) * E2

这个公式可以这样读:计算从单元格 B2 到单元格 D2 的数的平均值,接着用这个值乘

以单元格 E2 的数。我们从单元格 F3 到单元格 F11 复制单元格 F2 中的公式，结果会在 F 列显示。

我们来命名：对某一范围的单元集

输入单元格地址如 A1:A3 非常容易，但这不是要关注的内容。

如果我们真正面对的是一个非常大的工作表，有上百行、上百列以及上千个单元格，这时如果只是输入一个名称，就能代表一定范围内所有的单元格而不需要记住所有这些单元格的地址，这不是非常好吗？Excel 就能提供这样的功能，允许你对一个范围或者是单元集进行命名。

例如，在图 1A.11 中，如果你想计算员工第二个评估分的平均值，你可以用单元格地址 C2:C11，也可以对所想要范围的单元格给予一个名称，如 eval2 或者 Eval_2（名称中一定不要添加空格）。接着，使用 AVERAGE 函数计算这一组数据的平均数将会是

=AVERAGE(Eval_2)

而不一定必须是

=AVERAGE(C2:C11)

还能做得更好——你只需要单击鼠标将名称粘贴到公式或者函数中，甚至都不需要手动输入。下面是如何给一定范围的单元格分配名称的步骤。

1.选中你需要命名的单元格范围。

2.单击[公式]→[定义名称]选项框。

3.输入你想用的名称，如图 1A.12 所示（这里使用的是 Eval_2）。再一次强调，没有空格——Excel 在名称定义中不接受空格。

4.按[确定]键。

Eval_2　fx 76

	A	B	C	D	E	F
1	Name	Eval 1	Eval 2	Eval 3	Fudge Factor	Final Score
2	HY	67	76	34	1.10	64.90
3	NM	65	78	32	1.02	59.50
4	GG	45	98	34	1.23	72.57
5	DF	65	98	54	1.06	76.67
6	KH	76	78	58	1.03	72.79
7	RR	32	76	75	1.04	63.44
8	YR	45	77	54	1.05	61.60
9	HH	43	76	54	1.00	57.67
10	JU	34	54	33	1.01	40.74
11	WE	32	47	78	1.04	54.43

图 1A.12　对选中的一定范围的单元格集命名

使用定义的范围

一旦对某一范围的单元格定义了名称，你就可以使用这个名称来替代这个范围下的单元集。如果你记得所命的名字，你可以使用[名称管理器]（Name Manager）对话框的下拉菜单直接选定这个名称，如图 1A.13 所示。

下面我们使用定义的范围来计算所有经验系数的平均数。

1.单击单元格 F12 作为平均值所在单元格。

2.输入 =AVERAGE(。

3.单击[公式]→[用于公式]（Use in Formula），你就可以看到下拉菜单，如图 1A.14 所示。

图 1A.13　查看范围的名称

AVERAGE =average(

	A	B	C	D	E	F
1	Name	Eval 1	Eval 2	Eval 3	Fudge Factor	Final Score
2	HY	67	76	34	1.10	64.90
3	NM	65	78	32	1.02	59.50
4	GG	45	98	34	1.23	72.57
5	DF	65	98	54	1.06	76.67
6	KH	76	78	58	1.03	72.79
7	RR	32	76	75	1.04	63.44
8	YR	45	77	54	1.05	61.60
9	HH	43	76	54	1.00	57.67
10	JU	34	54	33	1.01	40.74
11	WE	32	47	78	1.04	54.43
12						=average(

图 1A.14 ［用于公式］下的菜单名录

4.单击 Fudge_Factor

5.输入)。

6.双击 Final Score,得到结果如图 1A.15 所示。

现在你已经了解了在公式中如何使用单元格范围的名称而不是单元格地址。

F12 =AVERAGE(Final_Score)

	A	B	C	D	E	F
1	Name	Eval 1	Eval 2	Eval 3	Fudge Factor	Final Score
2	HY	67	76	34	1.10	64.90
3	NM	65	78	32	1.02	59.50
4	GG	45	98	34	1.23	72.57
5	DF	65	98	54	1.06	76.67
6	KH	76	78	58	1.03	72.79
7	RR	32	76	75	1.04	63.44
8	YR	45	77	54	1.05	61.60
9	HH	43	76	54	1.00	57.67
10	JU	34	54	33	1.01	40.74
11	WE	32	47	78	1.04	54.43
12						62.43

图 1A.15 在函数中插入一个单元格范围的名称

小 结

这章内容很“少”,但是包含了 Excel 提供的一些最有用的工具。事实上,公式和函数的使用只是受到了你自己想象的局限。当你使用 Excel,你会找到更多的方式来让这些工具成为供你驱遣的强大的工具。

练 习

1.在一个 Excel 工作表中按要求建立公式。

a.3 加 5 的和。

b.10 减去 5,所得差乘以 7。

c.5、6、7 和 8 的平均数。

d.3、4 和 5 的平方求和。

2.使用 AVERAGE 函数计算 3、5 和 7 的和再除以 3。
3.计算单元格 A1 到 A5 的数值加和的函数。
4.计算单元格 A1 到 A5 的平均数的函数。
5.建立一个工作表，给单元格 A1 到 A5 任意赋值，定义这个单元格范围的名称为 Test_Scores，在单元格 A6 建立公式，先计算这些数的和，再计算平均值。你可以想到其他的完成方式不？

表 1A.1

你会越来越爱函数。要注意与 2007 版（以及之前版本）相比，Excel 中的许多函数发生了一些小的变化。例如 TTEST 现在是 T.TEST（当然给了广大的 Excel 用户许多线索去发现为什么发生变化）。

公式名称	公式含义	所在章节
AVERAGE	返回参数的平均数	2
GEOMEAN	返回几何平均数	2
MEDIAN	返回给定数字集的中位数	2
MODE.SNGL	返回一个数据集的众数（一个，默认为第一个）	2
MODE.MULTI	返回一个数据集的众数（可将大于 1 个的各个众数都返回）	2
QUARTILE.INC	返回一个数据集的四分位数	2
STDEV.S	估计样本标准差	3
STDEV.P	计算总体的标准差	3
VAR.S	估计样本的方差	3
VAR.P	计算总体的方差	3
KURT	返回一个数据集的峰度	4
SKEW	返回一个分布的偏度	4
CORREL	返回两个数据集之间的相关系数	5, 16
PEARSON	返回皮尔森积距相关系数	5, 16
NORM.S.DIST	返回标准正态累积分布	8
STANDARDIZE	返回标准化的值	8
Z.TEST	返回单侧 z 检验值的概率	10
T.DIST	返回学生 t 分布	11,12
T.TEST	返回学生 t 检验值的概率	11,12
F.DIST	返回 F 概率分布	13
F.TEST	返回 F 检验的结果	13
FORECAST	返回趋势线的一个数值	16
FREQUENCY	返回纵列呈现的频数分布	16
LINEST	返回趋势线的参数	16
SLOPE	返回线性回归线的斜率	16
STEYX	返回回归方程中依据 x 预测 y 的估计标准误	16
TREND	返回趋势线的数值	16
CHISQ.DIST	返回单侧卡方分布的概率	17
CHISQ.TEST	返回独立性检验	17

1B 使用奇妙的分析工具库应该了解什么

本章你会学到什么

- 分析工具库(Analysis ToolPak)是什么以及如何使用

你必需了解的几乎所有 Excel 的基本内容,都在附录 A 里。但是特定的 Excel 过程只有在安装了分析工具库(之前叫作数据分析工具包)之后才可以使用,而且我们会在本书很多章节使用这些工具。Excel 将这些工具集作为分析工具库,但是你在屏幕上看到的可能是 Data Analysis Tools(数据分析工具)或者 Data Analysis(数据分析),这些都不用担心。分析工具库是 Excel 的附加软件,你电脑中的 Excel 可能没有安装这个特定的工具集。

如何知道你的电脑是否安装了这个软件? 如果数据分析(Data Analyasis)选项没有在数据栏(Data Tab)(一般在数据栏的右侧)出现,你就需要安装分析工具库。

分析工具库非常好用。你只要按照步骤,明确你要做什么分析以及分析什么数据,就可以使用了。在本书我们会展示如何完成分析的详细步骤。

还有就是,你已经通过表 1A.1 了解到,公式非常多,其中一些公式在分析工具库中具有对应的公式。例如,你可以使用一个函数(T.TEST)检验两个群体的均值的差异,也可以使用分析工具库中的 *t*-test 工具。两种方法都可以。你经常可以有多种方式来分析数据——总有一种适合你。

分析工具库初探

现在,我们来看看从 25 个数中选取 5 个数的随机抽样的结果。在图 1B.1 中,你可以看到所有的数在 A 列,5 个数的样本在 C 列;我们从分析工具库使用的抽样对话框"告诉" Excel 如何抽样,样本中几个数,结果放在哪儿。在本书后面的部分会介绍更多相关内容,现在我们认为你想了解这些工具集如何工作。

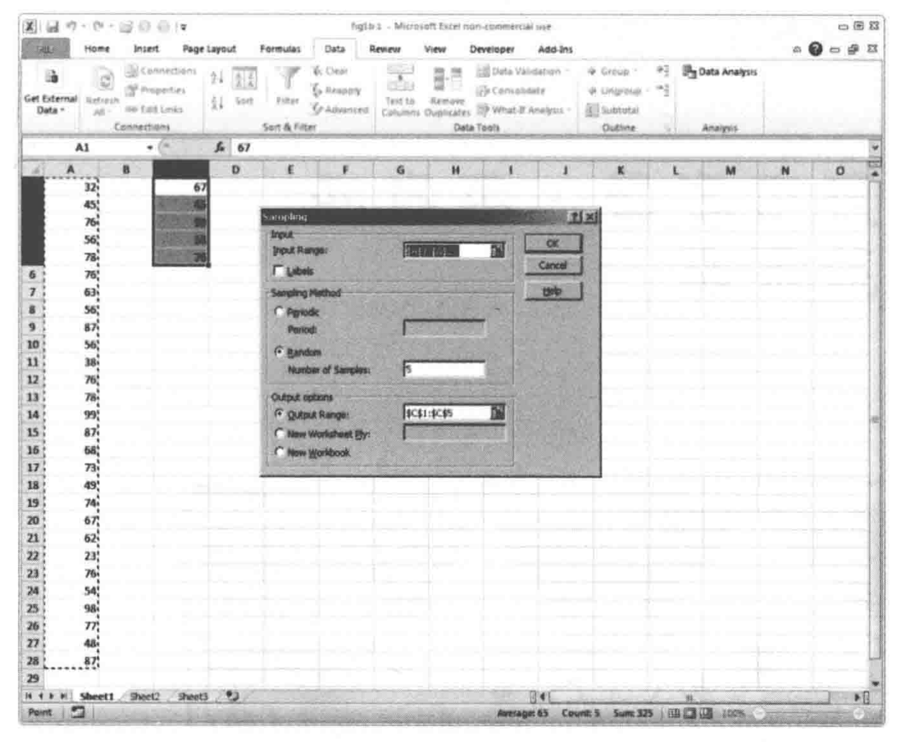


图 1B.1 使用分析工具库中的抽样工具

我们会继续介绍其他分析工具库中的工具(我们可能会介绍其中的四分之三,剩下的工具则超过了本书学习内容的需要)。在对应的章节(如第 2 章)我们会学习特定的工具(如描述统计)。

ANOVA(方差分析)
Correlation(相关分析)
Descriptive Statistics(描述统计)
Histogram(条形图)
Random Number Generation(随机数字生成器)
Rank and Percentile(排序和百分位)
Moving Average(移动平均数)
Sampling(抽样)
Regression(回归分析)
Z-test(*z* 检验)
t-test(*t* 检验)

没有分析工具库怎么办?

分析工具库是 Excel 的加载项。如果没有,可以按照下面的步骤在 Excel 2010 中安装分析工具库:

1.单击[文件]选项,接着单击[其他]选项。

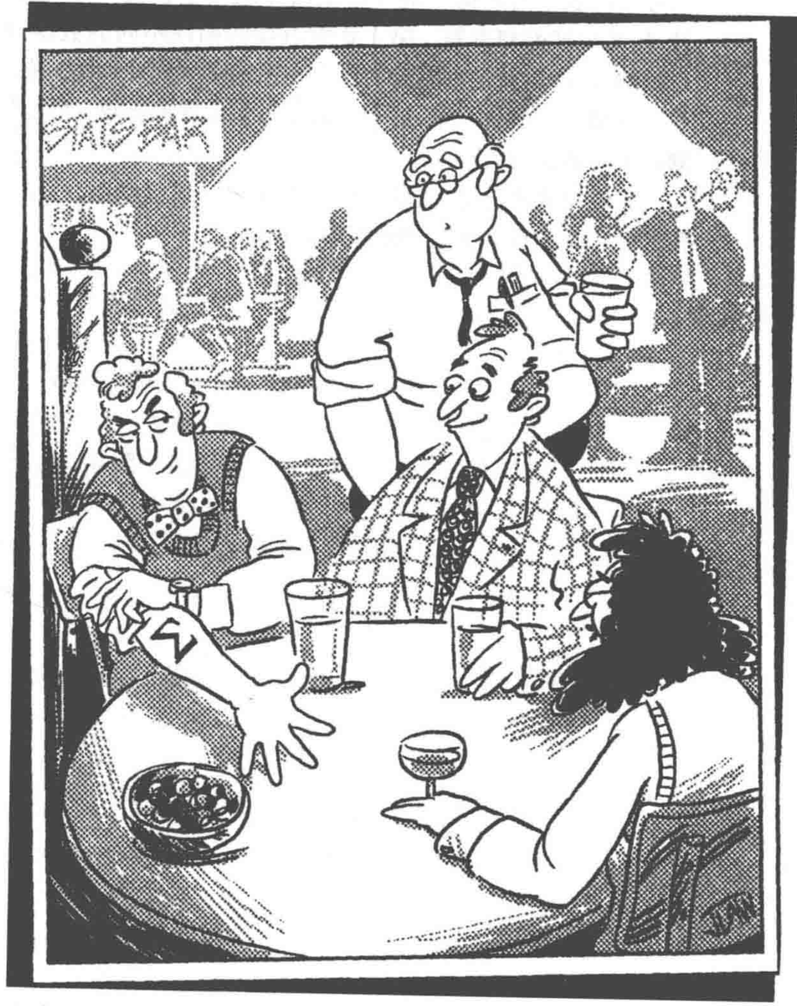
2.选择[加载项],接着在[管理对话框]选择[Excel 加载项]。

3.单击[转到]。

4.在[加载宏]对话框,选择[分析工具库]复选框,然后单击[确定]。

现在你已经完成安装,可以让你的 Excel 的运行更加富有成效并充满乐趣。

西格玛·弗洛伊德和描述统计



你觉得你的统计学教授很难搞?

精神分析学的奠基人西格玛·弗洛伊德(Sigmund Freud)* 做得很好的一件事就是观察和描述病人的状况。他是敏锐的观察者，应用他的技能发展了第一个系统而综合的人格理论。不论你对他的观点的有效性持什么看法，他都是一个优秀的科学家。

回到20世纪早期，统计学课程(就如你所选的统计课程)不是大学生或研究生课程表中的必修课。那个时候统计学还是相当新的领域，而且科学探索的精神尚不需要统计工具带给科学界的精确性。

但是形势已经发生改变。现在在任何领域，数据都很重要。《爱上统计学：Excel》的这部分主要致力于，在整理了所获数据之后，如何应用统计学方法描述结果并更好地理解结果。

第2章讨论集中趋势的测量，多种平均数如何计算以及何时用哪一种以便最好地代表一组数据。第3章全面介绍用于描述一系列数据点的离散性的工具，包括标准差和方差。第4章学习如何描述不同的数据分布或不同的数据集之间的差异，以及这种差异的意义。第5章处理变量之间的关系，也就是相关关系。这部分的最后一章，第6章的重点是信度和效度，这两者是评估所有研究中测量工具的关键。

完成第Ⅱ部分之后，你已经站在了很好的起点上，可以开始理解概率和推论在社会科学和行为科学中扮演的角色。

---

* Sigmund Freud，一般译为西格蒙德·弗洛伊德。此处作者为了使语言更诙谐有趣，所以采用了“Σ”符号来代替 Sig，故此处将其译为“西格玛”。——译者注

必须完成的功课——计算和理解平均数 2

本章你会学到什么

- 了解集中趋势量数
- 使用 AVERAGE 函数计算一组数据的均值
- 使用 MODE 函数计算一组数据的众数
- 使用 MEDIAN 函数计算一组数据的中位数
- 使用分析工具库计算描述统计值
- 选择一种集中趋势量数

能看到这里你已经很有耐心了，那么现在就开始处理真实、生动的数据吧！这也正是你在这一章要做的。一旦收集了数据，第一步通常是整理数据，也就是使用简单的指标来描述数据。完成这一步最容易的方法就是计算几种不同形式的平均数。

平均数(average)是最能够代表一组数据的数值。数据组可以是 30 个五年级学生的拼写测验的成绩，或是纽约洋基队的击打率，或是最近的选举中登记为民主党或共和党的人数，这不重要。在所有的案例中数据组都可以使用平均数来概括。平均数也叫作**集中趋势量数**(measures of central tendency)，一般有三种形式：均值、中位数和众数。每一种形式提供不同的数据分布信息，而且计算和解释都很简单。

计算均值

均值(mean)是计算平均数最常用的形式。均值很简单，就是数据组中所有数值的总和除以该组数值的个数。如果你有 30 个五年级学生的拼写成绩，将所有的拼写成绩简单地加起来就得到一个总和，然后除以学生的人数，也就是 30，这样就得到均值。

公式 2.1 就是均值计算公式。

$$\bar{X} = \frac{\sum X}{n} \tag{2.1}$$

其中：

- 上带横线的字母 X(读作"X 拔")是数据组的平均数或均值；
- $\sum$ 或希腊字母西格玛是连加符号，也就是将其后的所有数值都加起来；
- X 是数据组中每个具体的数值；
- n 是从中计算均值的样本的规模。

下面是计算均值的步骤：

1—以一列或多列的形式列出所有数值，这些都是 X。

2—计算所有数值的总和。

3—总和除以数值的个数。

例如，你要计算三个不同商店消费者数量的平均数，就可以计算均值。

场　所	年顾客数量
兰哈姆公园商店	2 150
威廉斯堡商店	1 534
城区商店	3 564

每个商店顾客数量的平均数或均值是 2 416。公式 2.2 表明如何应用公式 2.1 计算这个值。

$$\bar{X}=\frac{\sum X}{n}=\frac{2\ 150+1\ 534+3\ 564}{3}=2\ 416 \tag{2.2}$$

瞧，我告诉过你很容易，小菜一碟！

现在，使用 Excel 的平均数函数

按照下面的步骤，使用 Excel 计算一组数据的均值。

由于某些原因，一些人宁愿将计算均值(mean)的函数叫作 AVERAGE，而不愿意叫作 MEAN。奇怪的是，同样是这些人将计算中位数(median)的函数命名为 MEDIAN，计算众数(mode)的函数命名为 MODE，那为什么不将计算平均数的函数命名为 MEAN，这样对大家来说都更简单？如果你找到了原因，请让我们知道。

1.在一个工作表的一列中输入数据，如图 2.1 所示。

2.选择你要输入 AVERAGE 函数的单元格。在这个案例中，我们在单元格 A5 计算均值。

3.我们可在 A5 单元格输入计算这三个数的平均数的公式，形式如下。

	A
1	2,150
2	1,534
3	3,564

图 2.1　用于计算均值的数据

=(A1+A2+A3)/3

或者单击单元格 A5，输入如下形式的 AVERAGE 函数(我们这样做了)。

=AVERAGE(A1:A3)

然后按回车键。

或者使用[公式]→[插入函数]菜单选项和我们之前在 1A 部分讨论的“插入一个函数”的技术，在单元格 A5 输入 AVERAGE 函数。



不论是输入一个函数还是使用[公式]→[插入函数]菜单选项,函数的形式都一样,一旦这个公式在那儿了,不论是输入还是插入,都是一回事儿。

如图 2.2 所示,均值已经计算,所得数值出现在单元格 A5。需要注意的是,在图 2.2 的公式栏(可以看到单元格的内容)中你可以看到 AVERAGE 函数的完整表达,计算的结果是 2 416,与我们手动计算的一样。

A5 =AVERAGE(A1:A3)

	A	B	C	D	E
1	2,150				
2	1,534				
3	3,564				
4					
5	2,416				

图 2.2　使用 AVERAGE 函数计算一组数据的均值

更多 Excel

你可能需要了解几何平均数(函数是 GEOMEAN)。几何平均数是在你有两个或者更多的度量单位来描述不同的数据集时使用。例如,你可能想要对医疗保健项目的顾客满意度按照 1~10 分来进行排序,但是对相同项目的医疗保健结果的质量按照 1~100 分进行排序。

需要记忆的内容



均值有时也用字母 M 表示,也叫作典型平均数或中心值。如果你在看其他的统计学读物或者研究报告时,看到类似 $M=45.87$ 这样的表达式,它的意思大概就是均值等于 45.87。

- 在上面的公式中,小写字母 n 表示从中计算均值的样本的规模。大写字母 N(像这样的)表示总体规模。在一些书中和一些杂志的文章中没有对这两者做出区别。
- 样本均值是能准确反映总体均值的集中趋势量数。
- 最后,谈不上是好是坏,均值对极值很敏感。极值会使得均值向一方或另一方倾斜,也使得均值对数据组的代表性减弱,同时作为集中趋势量数的有效性减弱。当然,均值对一组数据的代表性高低取决于这组数据的数值,后面会有更多的讨论。

Tech Talk

均值也指**算术平均数**(arithmetic mean),当然你也会看到其他类型的均值,如调和均值。调和均值用于特殊的情况,你现在不需要关注。如果你想表现得更专业,就需要知道算术平均数(也就是我们讨论到现在的均值)也被定义为数据集中所有数值与其的差之总和为 0 的那个数值。例如你有三个值 3、4 和 5(均值是 4),均值的偏差(-1、0 和 1)的总和是 0(耶!)。

更多 Excel

除了算数平均数(许多平均数之一)之外,分析工具库中还有移动平均(Moving Average)。使用移动平均工具,可以计算大量数据的平均数。例如,你现有数据 1、4、5 和 10。这些数的平均数是 5。移动平均数每次只计算两个数的平均数(1+4 的平均数,4+5 的平均数,5+10 的平均数),将这些平均数再次平均得出的数是 4.88。由于移动平均数考虑了数据中的极值或奇异值(在这个数据集中是 10),因此在一些情况下移动平均数相对来说更准确。移动平均工具也可以绘制图表来图示每一数据点的真实数据而不是平均数。

记住“平均数”一词仅仅是最能代表一组数据的测量值之一,而且还有许多不同类型的平均数。使用哪一类型的平均数取决于你所提的问题和你要汇总分析的数据的类型。

计算加权平均数

以上的案例是如何计算简单的均值。但是有的情况下同样的数值不止出现一次,这就需要计算加权平均数。加权平均数很容易计算,也就是每一数值乘以它出现的频数,并将所有的积相加,然后除以频数的总和。



按照下面的步骤计算加权平均数:

1—列出要计算均值的样本的所有数值,如下表中“数值”列中的所有数值(X 的值)。

2—列出每一个数值出现的频数。

3—每一个数值乘以它的频数,如表中第三列所示。

4—计算“数值×频数”列的所有数值总和。

5—除以频数的总和。

例如,这个表格整理的数据是 100 个飞行员飞行熟练度测验的数值和频数。

数　值	频　数	数值×频数
97	4	388
94	11	1 034
92	12	1 104
91	21	1 911
90	30	2 700
89	12	1 068
78	9	702
60(不要和这个家伙一起飞行)	1	60
总计	100	8 967

加权平均数是 8 967/100,或 89.67。以这种方式计算均值相对于将 100 个不同的值输

入计算器或者计算机软件要容易得多。

我们如何使用 Excel 计算加权平均数？非常容易。按照下面的步骤做。

1.在新的工作表中输入上表数据。

2.建立一个公式将数值与对应的频数相乘，公式见单元格 C2，如图 2.3 所示。

	A	B	C
1	Value	Frequency	Value x Frequency
2	97	4	=A2*B2
3	94	11	
4	92	12	
5	91	21	
6	90	30	
7	89	12	
8	78	9	
9	60	1	

图 2.3　使用 Excel 计算加权平均数

C15

	A	B	C
1	Value	Frequency	Value x Frequency
2	97	4	388
3	94	11	1,034
4	92	12	1,104
5	91	21	1,911
6	90	30	2,700
7	89	12	1,068
8	78	9	702
9	60	1	60
10		100	8,967
11			
12		Weighted Average	89.67

图 2.4　计算加权平均数

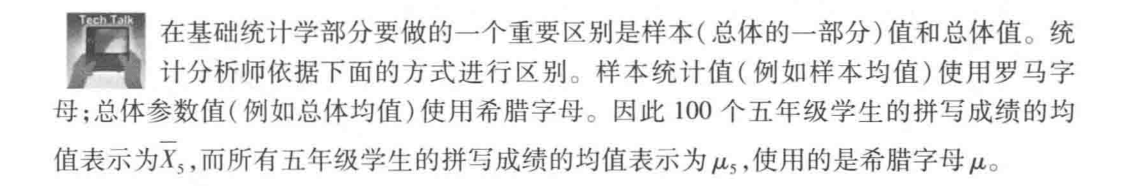
3.在这一列复制这个公式，这样 C2 到 C9 的数值就是乘积。

更多 Excel

记住一个单元格可以包含的信息。有时可能是公式，即使我们看到是这个公式的结果。如图 2.3 中，我们使用 Ctrl+` 综合键来显示单元格中的公式或者函数，而不是公式或者函数的结果。

4.在单元格 B10 和 C10，使用 SUM 函数来计算所对应列的和。

5.现在，将结果放在单元格 C12，总和(单元格 C10)除以总频数(单元格 B10)，如图2.4所示。哈哈！你又做了一次，得到的加权平均数是 89.67，与前面的计算结果一样。



在基础统计学部分要做的一个重要区别是样本(总体的一部分)值和总体值。统计分析师依据下面的方式进行区别。样本统计值(例如样本均值)使用罗马字母；总体参数值(例如总体均值)使用希腊字母。因此 100 个五年级学生的拼写成绩的均值表示为$\overline{X}_5$，而所有五年级学生的拼写成绩的均值表示为μ_5，使用的是希腊字母μ。

计算中位数

中位数也是平均数，却是一种非常不同的形式。中位数(median)被定义为一系列数据的中点。在这一数据点，所有数据的一半也就是 50%在其之上，而另一半或者说 50%在其之下。中位数有一些特别的性质，我们会在这一节后面的部分讨论，现在集中精力考虑如何计算。计算中位数没有标准公式。



下面是计算中位数的步骤：

1—以从大到小或者从小到大的顺序列出数据；

2—找到位于中间位置的数值，那就是中位数。

例如，下面是 5 个家庭的收入：

135 456 美元

25 500 美元

32 456 美元

54 365 美元

37 668 美元

下面是将收入从大到小排序后的清单：

135 456 美元

54 365 美元

37 668 美元

32 456 美元

25 500 美元

一共是 5 个数。中间位置的数是 37 668 美元，它就是中位数。

现在，如果数据的个数是偶数怎么计算？我们在收入清单中增加一个数据（34 500 美元），那么就有 6 个收入数据，如下所示：

135 456 美元

54 365 美元

37 668 美元

34 500 美元

32 456 美元

25 500 美元

如果数据个数是偶数，中位数就是中间两个数值的平均数。在这个案例中，中间的两个数是 34 500 美元和 37 668 美元，这两个数的平均数是36 084美元。这就是 6 个数据的数据集的中位数。

如果中间的两个数相同，怎么计算？就如下面的数据集：

45 678 美元

25 567 美元

25 567 美元

13 234 美元

那么中位数和中间的两个数相同。在这个案例中中位数是 25 567 美元。

现在，使用 Excel 的中位数函数

按照下面的步骤，使用 Excel 计算一组数据的中位数。

1.在工作表中将上面的六个数输入一列，如图 2.5 所示。

2.选择你要输入 MEDIAN 函数的单元格。在这个案例中，我们在单元格 A9 计算中位数。

3.单击单元格 A9，输入中位数函数（如下所示）然后按回车键。

= MEDIAN(A2:A7)

或者使用[公式]→[插入函数]菜单选项和我们在 1A 部分讨论的“插入函数”技术，输入 MEDIAN 函数。

在图 2.6 中，你可以看到 A9 单元格中显示了中位数，同时，在公式栏你可以看到 MEDIAN 函数。

	A
1	Income Level
2	$135,456
3	$54,365
4	$37,668
5	$34,500
6	$32,456
7	$25,500

图 2.5　计算中位数的数据

A9　=MEDIAN(A2:A7)

	A	B	C	D	E
1	Income Level				
2	$135,456				
3	$54,365				
4	$37,668				
5	$34,500				
6	$32,456				
7	$25,500				
8					
9	$36,084				

图 2.6　使用 MEDIAN 函数计算中位数

如果你了解中位数，也应该知道百分位点(percentile point)。百分位点用于定义数据集或数据分布中等于或者小于一个特定数据值的个体的百分数。例如，你的成绩是“处在 75 百分位点”，这意味着成绩分布中你的成绩刚好处于或者超过 75%的其他人的成绩。大家也知道中位数是 50 百分位点，因为数据分布中 50%的个体在这一点之下。其他的百分位点也很有用，如 25 百分位点通常用 Q_1 表述，75 百分位点，通常用 Q_3 表示。那么 Q_2 呢？当然是中位数。

现在给出一些问题的答案，这些问题可能在开始讨论中位数时就出现在你的头脑中了。为什么有时候要使用中位数而不是均值？主要原因是中位数对极值不敏感而均值却相反。

如果你的数据集中有一个或多个极值，中位数相对其他集中趋势量数来说能更好地代表数据集的中心值。是的，甚至比均值更好。

我们所说的极值是什么？非常容易就可以想到，极值就是与其所属数据组非常不同的值。例如，考虑之前案例中已经用过的收入清单，在这里再次列出：

135 456 美元

54 365 美元

37 668 美元

32 456 美元

25 500 美元

这组数据中数值 135 456 美元与其他 4 个数之间的差异较大。我们可以认为这个数值就是极值。

最好的说明中位数作为集中趋势量数是多么有用的方法，是计算包含一个或多个极值的同一数据集的均值和中位数，然后比较哪一个值能更好地代表数据集。现在就来计算和比较。

上面有 5 个收入数值的数据组的均值是 5 个数值的总和除以 5，结果是 57 089 美元。而这个数据组的中位数是37 668美元。哪一个值能更好地代表这个数据组？数值 37 668 美元更加明显地位于数据组的中间，但是我们习惯于认为平均数具有代表性或者占据中间位置。事实上，均值 57 089 美元在第 4 高位数值(54 365 美元)之上，不是这个数据分布

的中间值，也不具有代表性。

就是由于这个原因，特定的社会和经济指标（大多数与收入相关）的分析使用中位数作为集中趋势量数，例如“美国家庭平均收入的中位数是……”，而不是使用均值来概括收入。存在太多极值改变或者明显地扭曲一个数据组或者数据分布的中心点的情况。

你已经知道均值有时用大写字母 *M* 表示而不是 $\overline{X}$，中位数也有表示符号。我喜欢使用字母 *M*，但是一些人会将 *M* 和均值混淆，因此他们使用 *Med* 或 *Mdn* 表示中位数。不要让这些绊住你的脚——只要记住什么是中位数以及中位数代表什么，你就不会存在适应这些不同符号的困难。

更多 Excel

你可以使用 QUARTILE 计算数据分布中的 25 百分位数、50 百分位数和 75 百分位数，以及其他百分位数。也可以用来算中位数喔，因为中位数就是 50 百分位数。

需要记忆的内容



这些有关中位数的内容十分重要也很有趣，需要记忆。

- 均值是一系列数值的中间点，而中位数是一系列个体的中间点。
- 因为中位数关注的是有多少个体而不是这些个体的数值，极值（有时也叫作奇异值）就不会产生影响。

计算众数

我们要学习的第三个也是最后一个集中趋势量数是众数，也是最笼统、最不精确的集中趋势量数，但是在理解特定的数据集的特征中扮演着非常重要的角色。众数（Mode）就是出现次数最多的数值。没有计算众数的公式。



按照下面的步骤计算众数：

1—列出一个数据分布中的所有数值，但是每一个数值只列出一次。

2—计算每个数值出现的次数。

3—出现次数最多的数值就是众数。

例如，调查 300 个人的政党背景会形成如下的数据分布结果。

政党背景	次数或频数
民主党	90
共和党	70
无党派人士	140

众数是出现次数最多的数值，在上面的例子中是无党派人士。这就是数据分布的众数。

现在,使用 Excel 的 MODE.SNGL 函数

按照下面的步骤,使用 Excel 计算一组数据的众数。

1.在一个工作表的一列中输入数据,如图 2.7 所示。这一列表示政党背景。要记住我们必须输入数字(而不是字符),这样 Excel 才可以计算数值出现的次数。我们也在相同的工作表中记录了编码,这样你可以一直记得每个数值代表哪个政党。

2.选择你要输入 MODE.SNGL 函数的单元格。在这个例子中,我们在单元格 B21 计算众数。

3.单击单元格 B21,输入 MODE.SNGL 函数,如下所示:

=MODE.SNGL(A2:A20)

然后按回车键。

或者使用[公式]→[插入函数]菜单选项和我们在 1A 部分讨论的“插入函数”技术在单元格 B21 输入 MODE 函数。你可以在图 2.8 中看到众数和公式栏中的函数。

	A	B
1	Party	
2	1	1 = Democrats
3	2	2 = Republicans
4	3	3 = Independents
5	3	
6	3	
7	2	
8	2	
9	1	
10	1	
11	1	
12	1	
13	1	
14	2	
15	2	
16	2	
17	2	
18	1	
19	1	
20	1	

图 2.7 计算众数的数据

B21 =MODE.SNGL(A2:A20)

	A	B	C
1	Party		
2	1	1 = Democrats	
3	2	2 = Republicans	
4	3	3 = Independents	
5	3		
6	3		
7	2		
8	2		
9	1		
10	1		
11	1		
12	1		
13	1		
14	2		
15	2		
16	2		
17	2		
18	1		
19	1		
20	1		
21	Mode	1	
22			

图 2.8 使用 MODE.SNGL 函数

> 你可以使用 COUNTIF 函数计算类别(字符)出现次数,这是很简单的计算众数的方式,且不必使用 MODE 函数。只要建立一列数据,接着使用这项功能,定义包含的单元格范围和数值。在我们的案例中,可以用 = COUNTIF(A2:A20,Democrat)*,这个函数会对所有 Democrat 的出现进行计数,而且结果会回到你选定的单元格。接着,你选择最大的值作为众数。

你知道计算众数时最容易、最经常出现的错误是什么吗?就是选择某个分类选项出现的次数而不是分类选项的标签本身。对一些人来说很容易就可以得出众数是 140 而不

* 使用这个函数时,数据录入可直接输入文本字符,如 Democrat,而不必转换成数字。——译者注

是无党派人士，为什么？因为他们看到的是数值出现的次数，而不是最经常出现的那个数值！这是一个稍不留神就会犯下的错误，因此你计算众数时一定要注意。

双峰分布的苹果派

如果数据分布中每一个数值的出现次数都相同，那么就没有众数。但是如果不止一个数值的出现频数相同，那么这个数据分布是多峰分布。数据集可能是双峰分布（有两个众数）的，如下面头发颜色构成的数据组。

头发颜色	次数或频数
红色	3
金色	5
黑色	8
棕色	8

在上面的案例中，数据是双峰分布，因为黑色头发和棕色头发出现的频数相同。如果众数相当接近但不是完全相同，数据分布也是双峰分布，如 10 个人的头发是黑色，9 个人的头发是棕色。问题就变成一种类别在多大的程度上与其他类别相区分？你的数据有三峰分布的吗？也就是说出现了三个频数相同的数值。一般来说是不会有的，特别是在处理大的数据集过程中，但出现的可能性是存在的。

现在，使用 MODE.MULT 函数

大家知道吗？Excel 2010 建立了可让我们计算多个众数的函数（之前的 MODE 函数只能处理单一众数的情况，所以称作 MODE.SNGL，这是你刚学习的函数）。这个函数很“狡猾”，一定要特别小心。

按照下面的步骤，使用 Excel 计算一组数据的多个众数。

1.在工作表的一列输入数据，如图 2.9 所示。这一列数据代表头发颜色。

2.选择你要输入 MODE.MULT 函数的单元格。在这个例子中，我们选择单元格 B26 到 B29 显示众数。我们选择四个单元格，以防这组数据有多个众数。一般会预估众数的个数，多选择几个单元格数。

3.输入 MODE 函数，如下所示

= MODE.MULT（A2:A25）

然后同时按 Ctrl+Shift+Enter 组合键。你预计会出现多个众数，就要这样做。

或者使用[公式]→[插入函数]菜单选项和我们在 1A 中讨论的“插入函数”技术在单元格 B26 到 B29 输入 MODE.MULT 函数。你可以在图 2.10 中看到公式栏中的函数，函数表达式中使用了大括号，这是为了创建一组数据。因为这组数据只有两个众数，所以在单元格 B28 和单元格 B29 显示的是 N/A；Excel“期望”有更多众数出现，但是并没有。

更多 Excel

我们使用特定的功能键 Ctrl+Shift+Enter 组合，因为我们选定多个单元格建立一个列（或者说一个数组）来显示计算结果。如图 2.10 所示，通过公式栏可看到，输入的公式使用了大括号，因为预知结果可能是一个数据组。

	A	B
1	Hair Color	
2	1	1 = Red
3	1	2 = Blond
4	1	3 = Black
5	2	4 = Brown
6	2	
7	2	
8	2	
9	2	
10	3	
11	3	
12	3	
13	3	
14	3	
15	3	
16	3	
17	3	
18	4	
19	4	
20	4	
21	4	
22	4	
23	4	
24	4	
25	4	

图 2.9 计算多个众数

B26 fx {=MODE.MULT(A2:A25)}

	A	B	C
3	1	2 = Blond	
4	1	3 = Black	
5	2	4=Brown	
6	2		
7	2		
8	2		
9	2		
10	3		
11	3		
12	3		
13	3		
14	3		
15	3		
16	3		
17	3		
18	4		
19	4		
20	4		
21	4		
22	4		
23	4		
24	4		
25	4		
26	Modes	3	
27		4	
28		#N/A	
29		#N/A	

图 2.10 使用 MODE.MULT 函数

使用分析工具库计算描述统计

这是我们第一次使用奇妙的分析工具库，这个工具库我们在 1B 部分作了介绍。

分析工具库中对应的特定工具就叫作描述统计(Descriptive Statistics)，可以一次计算多个数值。你不能选择工具库计算什么(但是就如你即将看到的，你可以编辑结果)，我们会向你展示所有结果，但是只讨论本章覆盖的内容。我们在之后的章节会重复相同的过程。

按照如下的步骤使用工具库计算描述统计。我们使用图 2.11 所示的数据，也是图2.5 所示的数据。

1.单击[数据]→[数据分析]，你就会看到如图2.12所示[数据分析](Data Analysis)对话框。

	A
1	Income Level
2	$135,456
3	$54,365
4	$37,668
5	$34,500
6	$32,456
7	$25,500

图 2.11 描述统计工具选项使用的数据

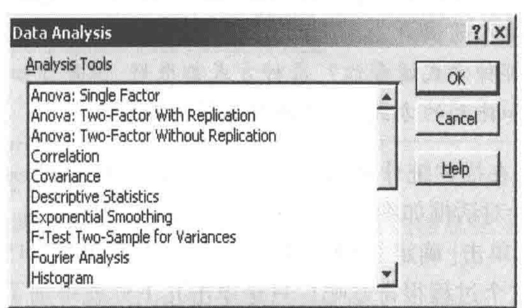


图 2.12 让我们开始使用分析工具库的对话框

2.单击[描述统计](Descriptive Statistics)，接着单击[确定](OK)，你就会看到如图2.13所示的对话框。

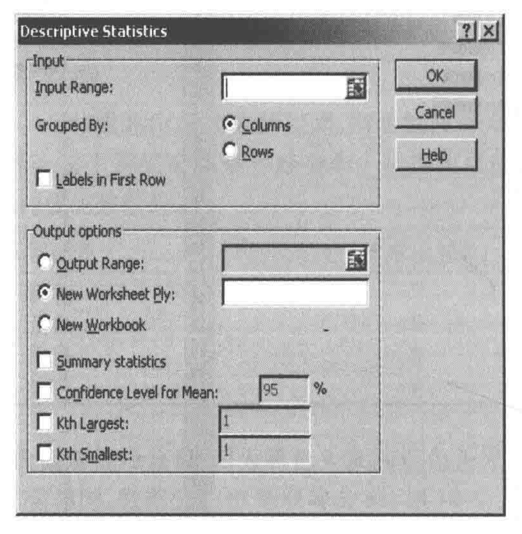


图 2.13　描述统计对话框

3.在[输入范围](Input Range)框中输入你想让 Excel 计算描述统计的数据的范围，但是要包括列的标题(这样的话列标题也会在分析中出现)。在这个例子中(如图 2.11 所示)，我们想分析的数据是从单元格 A1 到 A7。

4.单击[标志位于第一行](Labels in First Row)选项框。

5.现在，在对话框[输出选项](Output Options)单击[输出区域](Output Range)选项，输入你想让 Excel 将分析结果返回的具体单元格位置。在这个例子中，我们选择 C1。

更多 Excel

单击还是拖动？

相关内容在附录 A，但是这个内容太重要了，需要在这里也强调。你可以在 Excel 中以三种方式输入单元格地址。

- 你可以使用键盘输入(例如输入 A1)。
- 你可以在你选择的单元格上单击和拖动鼠标，接着放开，单元格范围就会在范围框中出现。
- 你可以使用[压缩](Collapse)按钮(就是▣)。单击之后，你也可以通过拖动鼠标方式输入范围。

哪种方式适合你？每种方式都很好，但是你知道输入方式有时候很费时。试着使用单击和拖动的方式，或者使用压缩按钮。

6.在描述统计对话框中单击[汇总统计](Summary Statistics)选项框。选择完成的描述统计对话框如图 2.14 所示。

7.单击[确定](OK)，接着你就会看到如图 2.15 所示的结果。

这个过程很奇妙吧！只是单击几下你就得到了所有的信息。你只需要确保将所有的准备工作做好，但是，必须保证所有的单元格都输入正确。

Descriptive Statistics
Input
Input Range: $A$1:$A$7
Grouped By: Columns / Rows
Labels in first row
Output options
Output Range: $C$1
New Worksheet Ply:
New Workbook
Summary statistics
Confidence Level for Mean: 95 %
Kth Largest: 1
Kth Smallest: 1
OK
Cancel
Help

图 2.14 选择完成的描述统计对话框

	A	B	C	D
1	Income Level		Income Level	
2	$135,456			
3	$54,365		Mean	53324.17
4	$37,668		Standard I	16887.72
5	$34,500		Median	36084
6	$32,456		Mode	#N/A
7	$25,500		Standard I	41366.29
8			Sample Va	1.71E+09
9			Kurtosis	4.861219
10			Skewness	2.173756
11			Range	109956
12			Minimum	25500
13			Maximum	135456
14			Sum	319945
15			Count	6

图 2.15 描述统计工具包的分析结果

你可以在图 2.15 看到所有有用的信息，包括这六个数的均值（实际是 53324.16667，四舍五入为 53324.17）、中位数（36084）和其他大量的信息（其中一部分本书不会讨论，你可以在高级统计学课程中学习）。

让分析工具库的输出结果更好看

一旦你使用分析工具库中的工具，并得到类似图 2.15 所示的输出结果，你当然就可以让输出结果保持那样，或者可以使用 Excel 工具改变输出形式以适应你的需要。输出结果绝对是建立工作表的一部分，因此你一开始在工作表做什么会对最后的输出结果产生影响。图 2.16 中你会看到我们已经使用简单的 Excel 工具做了简单的修改。

H28 fx

	A	B	C	D
1	Income Level		Income Level	
2	$135,456			
3	$54,365		Mean	53,324.17
4	$37,668		Standard Error	16,887.72
5	$34,500		Median	36,084.00
6	$32,456		Standard Deviation	41,366.29
7	$25,500		Sample Variance	1,711,170,163.37
8			Kurtosis	4.86
9			Skewness	2.17
10			Range	109,956
11			Minimum	25,500
12			Maximum	135,456
13			Sum	319,945
14			Count	6

图 2.16 新的、调整后的描述统计结果

- 我们将整个工作表的字体改为 Arial 12。
- 我们删除众数所在单元格，因为没有众数。
- 我们使用[格式]→[列]→[自动调整列宽]选项来调整列，让所有的信息在工作表中恰当地显示。
- 我们可以增加其他的形式，使用更多的 Excel 功能（颜色、表格形式、阴影等），但是你所见的形式已经可以很好地展示分析的结果。炫酷当然很好，但是简单和直接也不错。

何时用什么

好吧，我们已经定义了三种不同的集中趋势量数，而且每一种都给出了简单明白的案例。但是还有一个最重要的问题没有回答，即何时使用哪一种？

一般来说，使用哪一种集中趋势量数取决于你描述的数据类型。毫无疑问，定性数据、类别数据或定类数据（如种族群体、眼睛颜色、收入档次、选举次序以及邻里位置）的集中趋势只可以用众数来描述。

例如，你不能用中心值来描述哪种政治立场在一个团体中占有优势，也不能使用均值——难道你可以得出结论说每一个人是半个共和党？而300个人中几乎一半（140）是无党派人士似乎是描述这个变量的最好方式。一般来说，中位数和均值最适合于定量数据，如身高、具体收入数额（不是分类变量）、年龄、考试成绩、反应时间和一定程度上完成工作需要的时间。

公平地说，均值是比中位数更精确的测量值，中位数是比众数更精确的测量值。这意味着其他条件相同的情况下应优先使用均值，而且均值也的确是最常用的集中趋势量数。但是，我们的确会遇到一些均值不适合作为集中趋势量数的情况——例如我们收集的数据是类别数据或定类数据，例如头发颜色，那么我们使用众数。因此，这里的三项原则可能会有一定的帮助。但是要记住例外始终存在。

1.如果数据属性是分类的，而且数值只属于一种类型，例如头发颜色、政治背景、邻里位置和宗教，就使用众数。在这种情况下各个分类之间是互斥的。

2.如果数据中包含极值，而你不想扭曲平均数（按照均值计算）就使用中位数，例如你感兴趣的是收入。

3.最后，如果数据不包括极值也不是分类数据就使用均值，例如考试得分或游50码需要的时间。

小　结

不论你的统计技术掌握到何种程度，几乎都要从简单的描述开始——因此理解简单的集中趋势的概念很重要。在下一章，我们学习另一个重要的描述概念——变异性，也就是数值之间如何不同。

练　习

1.计算附录C的第2章数据集1的三个数据组（score）的均值、中位数和众数。手动计算，或者使用Excel计算。如果使用Excel，打印输出结果。

数据组1	数据组2	数据组3
3	34	154
7	54	167
5	17	132
4	26	145
5	34	154
6	25	145
7	14	113
8	24	156
6	25	154
5	23	123

2.使用分析工具库和“第2章数据集2”计算三个班的学生一周中每天提一个问题所需的平均时间。你使用哪一种集中趋势量数,为什么?

3.假定你是快餐店经理。你的工作之一是向老板汇报每天哪一种特惠食品卖得最好。应用你学到的描述统计的知识,写一个简短的报告让老板知道每天的经营状况。下面是数据,手动计算,记住备份你的计算结果。

特惠食品	售出数量	价格(美元)
Huge Burger	20	$ 2.95
Baby Burger	18	$ 1.49
Chicken Littles	25	$ 3.50
Porker Burger	19	$ 2.95
Yummy Burger	17	$ 1.99
Coney Dog	20	$ 1.99
总共销售	119	

4.在什么情况下你会使用中位数而不是均值作为集中趋势量数?为什么?给出两个例子说明在什么情形下中位数作为集中趋势量数比均值更有效。

5.列出三个变量,这三个变量如何测量可保证你使用众数是最适合的集中趋势量数。

6.假定你有自己的生意,你刚好购买并经营名为“Havefun.com”的网站,你通过网络销售某些玩具和某些游戏。现在评估你的第三季度(7月1日到9月31日)的广告预算,需要计算均值(mean)。(在Excel中是指“average”,使用名为AVERAGE的函数计算,还记得不?)下面是销售数据。使用Excel函数或者公式计算每个玩具、每月的销售额平均数。

玩具	7月销售额(美元)	8月销售额(美元)	9月销售额(美元)	销售均值
Slammer	12 345	14 453	15 435	
Radar Zinger	31 454	34 567	29 678	
Potato Gun	3 253	3 121	5 131	
销售均值				

7.使用第6题的数据,计算三种玩具各自的销售额中位数。

8.假设你处理的数据集中有一些数据非常“不一样”(比其他的数据大很多或者小很多)。你会使用什么样的集中趋势量数,为什么?

9.假定你是公共医疗部门的头,你想进行每周调查,了解在你的管理范围内不同年龄群体的感冒病例数。手动计算每周的均值和中位数。考虑数据的特殊性,你认为哪个指标是最有用的集中趋势量数。

	12/1~12/7	12/8~12/15	12/16~12/23
0~4岁	12	14	15
5~9岁	15	12	14
10~14岁	12	24	21
15~19岁	38	12	19
均值			
中位数			

3 差异万岁——理解变异性

本章你会学到什么

- 为什么变异性是有用的描述工具
- 如何计算极差、标准差和方差
- 标准差和方差如何类似,又如何不同
- 使用分析工具库计算极差、标准和方差

为什么理解变异性很重要

在第2章已经学习了不同类型的平均数,它们的含义,如何计算,以及何时使用。但是谈到描述统计和数据分布特征的描述,平均数只是一部分,另一部分是变异性(或离散)量数。

用简单的话说,**变异性**(Variability)反映数值之间的不同。例如下面的一组数据体现了一定程度的变异性:

7,6,3,3,1

接下来的另一组数据具有与之相同的均值(4),但变异性小一些:

3,4,4,5,4

下一个数据集根本没有变异性——数值之间没有差异,但是还是和前面所示的两组数据具有相同的均值。

4,4,4,4,4

变异性(也叫散布或离散度)可被看作是对不同数值之间的差异性的测量。其实把变异性看作是每个数值与特定值之间的差异程度可能更精确(而且也许更容易)。那么你认为哪一个“数值”可能是那个“特定值”呢? 通常情况下这个“特定值”就是均值——很正确。因此,变异性成为测量数据组中每一个数值与均值的差异程度的量数。接下来还会继续讨论。

要记住你已了解的关于计算平均数的内容——平均数(不论是均值、中位数还是众数)是一组数据的代表数值。那么现在增加关于变异性的新知识——它反映的是不同数值相对某一个值的差异性。两者都是重要的描述统计值。这两个值(平均数和变异性)可共同用于描述数据分布的特征,并说明数据分布之间的差异。

变异性的三种量数通常用于反映一组数据的变异性、散布或者离散度。这三种量数就是极差、标准差和方差。接下来让我们更详细地了解每一个量数及其如何应用。

计算极差

极差是对变异性最笼统的测量。极差可让你了解数值之间彼此差异的程度。**极差**(range)通过数据分布中的最大值减去最小值来计算。

一般来说,极差的计算公式如下:

$$r = h - l \tag{3.1}$$

其中:r 是极差,

h 是数据集中的最大值,

l 是数据集中的最小值。

以下面的数据组(以降序形式排列)为例:

98,86,77,56,48

在这个案例中,98−48=50。极差是 50。



实际上极差有两种类型:第一种是不包含极差,就是用最大值减去最小值($h-l$),也是我们刚刚定义的极差;第二种极差是包含极差,就是最大值减去最小值再加1,即($h-l+1$)。在研究性文章中通常看到的是不包含极差,但一些研究人员更喜欢使用包含极差,所以也偶尔会在文章中看到。

极差基本上都只用于得到不同值之间相互差别或离散程度的非常笼统的估计——极差仅表示一个数据分布中最小值到最大值之间的距离。

因此,虽然极差作为变异性的一般指标很好,但是不可以用于得出任何关于具体的数值之间相互差别的结论。

要记住,极差只使用了两个数值——是对一个数据集的变异性的非常表面的反映。

Excel 没有计算极差的函数,但是可以用分析工具库的描述统计选项来计算,就如第 2 章图 2.15 所示。另外,你可以建立简单的公式,一个数减去另一个数再加 1(计算包含极差)或者不加 1(计算不包含极差)。

计算标准差

现在开始学习最常用到的变异性量数——标准差。如果仅仅考虑字面的含义,就是标准化了的与某个值(猜猜是哪一个值)的偏差。实际上,标准差(standard deviation,缩写为 s 或 SD)表示一个数据组中变异性的平均数量。实际的含义是与均值的平均距离。标准差越大,每一个数据点与该数据分布的均值的平均距离越大。

那么,支持标准差计算的逻辑是什么?你最初的想法可能是计算数据组的均值,接着用均值减去每一个数值,然后计算这些距离的平均数。

这个想法很好——最后你会得到每一个数值与均值的平均距离。但是实际上这不管用(你是否知道为什么?接下来我们会说明原因的)。

下面给出计算标准差的公式：

$$s = \sqrt{\frac{\sum (X - \overline{X})^2}{n - 1}} \tag{3.2}$$

其中：s 是标准差，

$\sum$ 是西格玛，表示将其后所有数值累加求和，

X 是具体的数值，

$\overline{X}$是所有数值的均值，

n 是样本规模。

公式中可以看到先求每一个数值和均值之间的差$(X-\overline{X})$，这个差平方，并且计算所有平方值的总和。然后用平方和除以样本规模（减去 1），最后求平方根。就如你所看到的，也和我们早已指出的一样，标准差是每个数值与均值的偏差的平均数。

我们用下面的数据逐步解释如何计算标准差：

5，8，5，4，6，7，8，8，3，6



1—列出每一个数值，数值如何排序不重要。

2—计算数据组的均值。

3—每一个数值减去均值。

下面就是我们现在完成的工作，其中 $X-\overline{X}$表示每一个数值与所有数值的均值之间的差，均值是 6。

X	$\overline{X}$	$X-\overline{X}$
8	6	8−6=+2
8	6	8−6=+2
8	6	8−6=+2
7	6	7−6=+1
6	6	6−6=0
6	6	6−6=0
5	6	5−6=−1
5	6	5−6=−1
4	6	4−6=−2
3	6	3−6=−3

4—计算每一个差值的平方。计算结果在对应的$(X-\overline{X})^2$ 列。

X	$(X-\bar{X})$	$(X-\bar{X})^2$
8	+2	4
8	+2	4
8	+2	4
7	+1	1
6	0	0
6	0	0
5	−1	1
5	−1	1
4	−2	4
3	−3	9
合计	0	28

5—计算所有与均值的差的平方的总和。如在上表中看到的,总和是 28。

6—平方和除以 $n-1$,也就是 $10-1=9$,那么 $28/9=3.11$。

7—计算 3.11 的平方根,结果是 1.76(四舍五入之后)。这就是这 10 个数值的标准差。

现在,使用 Excel 的 STDEV.S 函数

按照下面的步骤,使用 Excel 计算一组数据的标准差(函数名称中的 S 表示样本)。

1.在工作表中的一列分数数据,如图 3.1 所示。

2.选择输入 STDEV.S 函数的单元格。在这个例子中,我们选择单元格 B12 计算标准差。

3.单击单元格 B12,输入 STDEV.S 函数,如下所示:

= STDEV.S(B2:B11)

然后按回车键。

或者使用[公式]→[插入函数]菜单选项和我们在 1A 中讨论的“插入函数”技术在单元格 B12 输入 STDEV.S 函数。

	A	B
1		Score
2		8
3		8
4		8
5		7
6		6
7		6
8		5
9		5
10		4
11		3

图 3.1 STDEV.S 函数所用数据

B12 =STDEV.S(B2:B11)

	A	B	C	D	E
1		Score			
2		8			
3		8			
4		8			
5		7			
6		6			
7		6			
8		5			
9		5			
10		4			
11		3			
12		1.76			

图 3.2 使用 STDEV.S 函数计算标准差

4.如图 3.2 所示,标准差已经计算出来。计算所得数值(1.76,与我们手动计算结果一样)返回到单元格 B12。要注意在图 3.2 所示的公式栏中,你可以看到标准差函数的完整表达。

更多 Excel

你喜欢什么公式？

这里会给大家一个惊喜。你在[公式]→[插入函数]选项中选择STDEV函数时，你会看到实际上有许多计算标准差的函数。其中的两个函数对我们非常重要，一个是STDEV.S函数（上面的例子中使用的函数），另一个是STDEV.P函数。这两个函数都计算数据组的标准差，但是它们的区别是什么？

我们用STDEV.S函数计算样本数据的标准差，STDEV.P计算总体数据的标准差。这两个数值之间有什么差别？为什么使用这个而不是另一个？之后会讨论。现在，因为STDEV.S函数是基于总体的一部分（样本的定义，对吧?），很有可能（我们将会讨论这个概率）高估总体的标准差——这个真的不能再真的真实值。

何时使用哪一个函数？使用STDEV.S函数几乎始终都是正确的。但是如果一组数据定义为总体数据，那么STDEV.P函数应该是你偏爱的函数。

从结果我们可以知道，数据分布中每一个数值与均值的偏差的平均数是1.76。

现在我们做一个简短的回顾，分析一下运用标准差公式的计算过程。这些过程很重要，需要回顾，而且会增加你对“标准差是什么”的理解。

首先，为什么我们不简单地将与均值的偏差累加起来？因为与均值偏差的总和始终都等于0。试着计算偏差(2+2+2+1+0+0-1-1-2-3)的总和。实际上，这也是检查均值计算是否正确的最好方法。



你可能会看到另一种类型的偏差，而且你应该了解它的含义。**平均偏差**（mean deviation，也叫作平均绝对偏差）是与均值偏差的绝对值的总和。你已经知道与均值偏差的总和一定等于0（否则就是均值计算错误）。那么，我们取每一个偏差的绝对值（也就是不管正负号的数值）。将这些绝对值累加起来，然后除以数据点的个数，你就得到平均偏差（注意：一个数的绝对值通常用两边带有竖线的数值表示，就如|5|。例如-6的绝对值是|-6|，也就是6）。

其次，为什么我们计算偏差的平方？因为我们想消除负号，这样就使得累加偏差的时候总和不为0。

最后，为什么第7步是以计算总值的平方根来结束？因为我们想回到开始时的计算单位。我们在第4步计算与均值偏差的平方（消除负值），接着在第7步计算总值的平方根，这就是有始有终。

O27

	A	B
1		Score
2		8
3		8
4		8
5		7
6		6
7		6
8		5
9		5
10		4
11		3
12	STDEV.S	1.76
13	STDEV.P	1.67

图3.3 STDEV.S **函数和** STDEV.P **函数比较**

为什么使用 $n-1$，而不是 n

你可能已经明白我们为什么对均值偏差进行平方，以及为什么求总和的平方根。但是为什么公式中的分母要减去1呢？为什么我们除以 $n-1$ 而不是 n 呢？这是个很好的问题。

答案是：s（标准差）是总体标准差的估计值，但是只有我们用 n 减去 1 的情况下才是**无偏估计**（unbiased estimate）。我们把分母减去 1 会使得标准差大于实际的大小。为什么我们要这样做？因为好的科学家一般都是保守的。这就是之前的例子中我们使用 STDEV.S 函数而不是 STDEV.P 函数的原因。实际上，看看图 3.3 你就会明白，在这个图中我们使用两个函数计算标准差。是否看出无偏估计（1.76）如何大于有偏估计（1.67）？那是因为基于样本的估计（无偏估计）是有目的地高估了这个数值。保守的意图是，如果我们不得不出错，也是由于过高估计了总体的标准差。除以较小的分母可让我们做到这一点。因此，我们除以 9 而不是 10，或者是除以 99 而不是 100。



如果你的目的是描述样本的特征，有偏估计也可以。但如果你想用样本估计总体参数，最好计算无偏统计值。

看看下面的数据表，你会发现样本规模变大（并且逐渐接近总体规模）会发生什么。$n-1$的调整对于标准差的有偏估计和无偏估计之间的差异的影响越来越小（见数据表中黑体字一栏）。其他各项一致的情况下，样本规模越大，标准差的有偏估计和无偏估计之间的差异越小。认真看下面的数据表，你会了解我们所说的内容。

样本规模	公式中分子的数值	总体标准差的有偏估计（除以 n）	总体标准差的无偏估计（除以 $n-1$）	有偏估计和无偏估计的差异
10	500	7.07	7.45	**0.38**
100	500	2.24	2.25	**0.01**
1 000	500	0.707 1	0.707 5	**0.000 4**

那么其中的规律是什么？当你计算样本的标准差用于估计总体的参数值时，样本规模越接近总体规模，估计就会越准确。

重要的是什么

标准差的计算很简单。但是标准差的含义是什么？作为一个变异性量数，标准差可以告诉我们数据组的每一个数值与均值的偏差平均数。但是就如将在第 4 章看到的，标准差有一些非常实际的应用。为了激发学习的兴趣，考虑这一点：标准差可用于帮助我们比较来自不同数据分布的数值，即使均值和标准差不同。很奇妙吧！这就如你将会看到的，非常酷。

需要记忆的内容



标准差是作为偏离均值的平均距离计算的。因此，你首先需要计算作为集中趋势量数的均值。因此计算标准差时不需要在中位数和众数上浪费时间。标准差越大，数值分布越广，则数值之间的相互差异越大。

和均值一样，标准差对极值很敏感。当你计算样本的标准差时，若数据中存在极值，你就要在数据报告中注明这一点。

如果 $s=0$，这组数据就没有一丁点儿变异性，而且各个数值都一样。这种情况很少发生。

计算方差

方差是另一个变异性量数，也是一个小惊喜。如果你知道一个数据组的标准差，并且计算其平方值，那就可以很容易地计算出这个数据组的方差。第三个变异性量数，也就是**方差**（variance），就是标准差的平方。

换句话说，就是你之前看到的公式，只是没有平方根符号，就如公式 3.3 所示：

$$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1} \tag{3.3}$$

如果你计算标准差时没有完成最后一步（计算平方根），那就得到方差。换句话说，$s^2 = s \times s$，即方差等于标准差自乘一次（或平方）。在我们的早前案例中，标准差等于1.76，方差等于 1.76^2，也就是 3.10。

现在，使用 Excel 的 VAR.S 函数

按照下面的步骤，使用 Excel 计算一组数据的方差。

1.在工作表中的一列输入数据。我们使用图 3.1 所示的相同的数据。

2.选择我们输入 VAR.S 函数的单元格。在这个例子中，我们在单元格 B12 计算方差。

3.现在单击单元格 B12，输入 VAR.S 函数，如下所示：

=VAR.S(B2:B11)

然后按回车键。

或者使用［公式］→［插入函数］命令菜单选项和我们 1A 中的“插入函数”技术在单元格 B12 输入 VAR.S 函数。

4.如图 3.4 所示，方差计算的结果返回到了单元格 B12。要注意你在图 3.4 的公式栏中可以看到 VAR.S 函数的完整表达式，而且计算所得数值是 3.11。这个数值与我们手动计算结果相同，唯一的不同就是 Excel 是使用内部公式和函数计算方差。

B12　　fx =VAR.S(B2:B11)

	A	B	C	D	E
1		Score			
2		8			
3		8			
4		8			
5		7			
6		6			
7		6			
8		5			
9		5			
10		4			
11		3			
12		3.11			

图 3.4　使用 VAR.S 函数计算方差

	A	B
1		Score
2		8
3		8
4		8
5		7
6		6
7		6
8		5
9		5
10		4
11		3
12	VAR.S	3.11
13	VAR.P	2.80

图 3.5　比较 VAR.S 函数和 VAR.P 函数

更多 Excel

正如 VAR.S 函数对应 VAR.P 函数,STDEV.S 函数对应 STDEV.P 函数。就如有计算总体标准差的函数(STDEV.P)一样,也有计算总体方差的函数,就是 VAR.P 函数。图 3.5中,你可以看到样本方差计算(VAR.S 函数)所得,大于它的“表亲”——VAR.P 函数计算所得的总体方差。

你不大可能在杂志的文章中看到方差被提到,或者看到方差被用作描述统计数据。这是因为方差值很难解释或应用于一组数据。毕竟,方差来自于偏差值的平方。

但是方差很重要,因为方差不仅是一个概念,也是许多统计公式和技术中实际应用的变异性量数。你会在本书后面的部分了解更多。

标准差与方差

标准差和方差哪些方面相同,哪些方面不同?

它们都是变异性、离散度或散布的量数。用于计算两者的公式非常类似。你常常会在学术论文的“结果”部分看到两者。

但是它们也非常不同。

首先,而且最重要的是,标准差(因为我们计算偏差平方和的平均数的平方根)以最初的计算单位存在。方差以平方单位存在(没有计算平方根)。

这是什么意思?假定我们需要知道一组装配电路板的生产工人的变异性。我们假定他们每个小时平均装配 8.6 个电路板,标准差是 1.59。数值 1.59 的意义是每个小时装配的电路板与均值的平均差异是 1.59 个电路板。

让我们看看方差的解释,方差是 1.59^2,或 2.53。方差可解释的含义是:工人每个小时安装的电路板与均值的平均差异的平方大约是 2.53 个电路板。这两个值哪个更有解释意义?

使用奇妙的分析工具库(再一次)

你知道吗?其实我们已经做过了。回顾第 2 章的图 2.12 到 2.15,展示了分析工具库运行描述统计分析,包含了标准差、方差和极差。这很简单,对不对?

小　结

变异性量数帮助我们更全面地了解数据点的分布。与集中趋势量数一起,我们可以使用这些数值来区别不同的数据分布,而且有效地描述一组考试分数、身高或个性测量得分的分布。现在我们可以思考和讨论数据分布,并且可以采用不同的方式来考察这些数据。

练　习

1.为什么极差是最方便的离散量数?而且是最不精确的变异性量数?我们何时使用极差?

2.计算下列项目的包含极差和不包含极差。

最高成绩	最低成绩	包含极差	不包含极差
7	6		
89	45		
34	14		
15	2		
1	1		

3.计算下面的数据集的极差、无偏标准差、有偏标准差和方差。请手动计算。

31,42,35,55,54,34,25,44,35

4.下面的数据是考试成绩，计算无偏标准差和方差。（提示：使用 STDEV.S 函数和 VAR.S 函数。）

87,87,58,78,98,69,88,71,86,91

5.计算第 4 题的数据的有偏估计。

6.为什么第 4 题计算的无偏估计大于第 5 题计算的有偏估计？

7.使用附录 C 第 3 章数据集 1 的数据来练习。这个数据集有两个变量。使用 Excel 计算身高和体重的标准差和方差（包括无偏估计和有偏估计）。

变　量	定　义
身高	身高是多少英寸
体重	体重是多少磅

8.你是大男孩航空公司年轻有为的决策者，你真的想比较上周最后两天飞往密苏里堪萨斯城、华盛顿特区和罗德岛的普罗维登斯的早航班和晚航班乘客人数的情况。下面列出了数据。计算你明天向老板汇报需要的描述统计值，并进行总结。要准时完成！

	星期四	星期五	星期四	星期五	星期四	星期五
早航班	堪萨斯城	堪萨斯城	华盛顿	华盛顿	普罗维登斯	普罗维登斯
乘客数	258	251	303	312	166	176
晚航班	堪萨斯城	堪萨斯城	华盛顿	华盛顿	普罗维登斯	普罗维登斯
乘客数	312	331	321	331	210	274

9.你在进行儿童阅读干预效果的实验，并且想向你的同事报描述统计结果。一些儿童接受了阅读干预，一些没有接受。平均数和标准差可以告诉你怎样的结果？使用第 3 章数据集 2 完成这个练习题。

一幅图真的相当于千言万语 4

本章你会学到什么

- 为什么一幅图真的相当于千言万语
- 如何建立直方图和多边形图
- 使用分析工具库建立直方图
- 使用偏度函数和峰度函数
- 使用 Excel 建立图表
- 使用 Excel 调整图表
- 不同类型的统计表和它们的应用
- 什么是数据透视表以及如何使用

为什么要用图表说明数据

前面的两章已经学习了两种类型的描述统计——集中趋势量数和变异性量数。一个可以提供描述一组数据的最佳值(集中趋势),另一个测量数值之间彼此差异或不同的程度(变异性)。

我们之前没有做而现在要做的是考察看起来不同的数据分布中两类测量结果的差异程度。数字本身(例如$\bar{X}=10$,$s=3$)很重要,但是用图形直观地展示,是更有效地描述数据分布特征和数据集特征的方式。

因此,在这一章我们将学习如何用图形直观表示数值的分布,以及如何使用不同类型的图形来表示不同类型的数据。

最后,我们会介绍数据透视表,是 Excel 新的附加功能,可以让我们按照需要设置行和列的显示。

好图表的十个方面(少贪新,多练习)

无论你是手动还是使用计算机软件制图,这些实用的图形设计原则都适用。这里有十条原则可供参照,最好贴在你的案头。

1.减少图或表中无用的内容。“图表垃圾”(类似于“废话”)是指使用了计算机软件的每一个函数、图表和性能,由此制作的图表中密密麻麻地挤满了没用的信息。这类东西绝对是越少越好。

2.在开始制作正式的图表之前要制作草图。即使准备使用计算机软件制作图表也要使用制图纸。

3.直抒胸臆，别整花样——传达的信息要不多不少刚刚好。没有比堆积图表（附带太多的文字和花哨的特征）迷惑读者更坏的事情了；也没有什么比简单明了的图形更好了。

4.给所有的内容贴标签，不要留下让读者不理解的内容。

5.一个图表应该只传递一个观点。

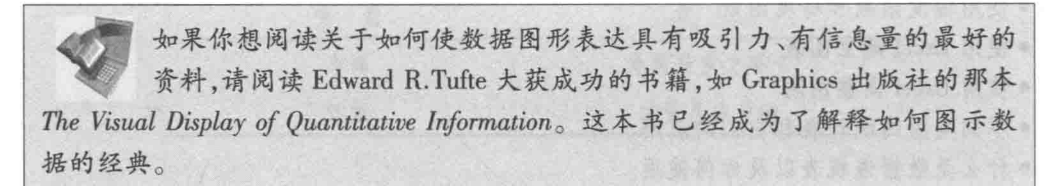
6.保持图表平衡。建立图表时，标题和数轴标签要居中。

7.保持图表的比例。比例是指横轴和纵轴之间的关系。这里的比例通常是 3∶4，也就是图表的宽若是 3 厘米，那么高就是 4 厘米。

8.简单最好。保持图表简单，但不能过于简化。尽可能只表达一个观点，而减去的信息在接下来的正文中保留。记住，一个图表应该是可以独立存在的，读者只看它就能够理解传递的信息。

9.限制你所使用的词语数目。词语太多或太长都会削弱你的图表传递的直观信息。

10.单凭图本身就要能够呈现你要表达的内容。如果不能，返回准备阶段重新制作。



如果你想阅读关于如何使数据图形表达具有吸引力、有信息量的最好的资料，请阅读 Edward R.Tufte 大获成功的书籍，如 Graphics 出版社的那本 *The Visual Display of Quantitative Information*。这本书已经成为了解释如何图示数据的经典。

初步之初：建立频数分布

用图表说明数据的最基本方式是建立频数分布。**频数分布**（frequency distribution）是记录和展现特定数据出现次数的方法。在建立频数分布过程中，数值通常依据一定的组距或数值范围分组。

下面是原始数据（50 个人阅读理解考试的成绩）和对应的频数分布表：

47	10	31	25	20
2	11	31	25	21
44	14	15	26	21
41	14	16	26	21
7	30	17	27	24
6	30	16	29	24
35	32	15	29	23
38	33	19	28	20
35	34	18	29	21
36	32	16	27	20

组　距	频　数
45~49	1
40~44	2
35~39	4
30~34	8
25~29	10
20~24	10
15~19	8
10~14	4
5~9	2
0~4	1

最合适的组距

就如你在上表中所看到的，**组距**（class interval）是一个值域范围，而且建立频数分布的第一步就是定义每一个组距的大小。在我们建立的频数分布中可以看到，每一个组距包含 5

个值，例如5~9（包含5、6、7、8、9）和40~44（包含40、41、42、43、44）。我们如何决定一个组距只包含5个值？为什么不是各包含10个数值的5个组距？或者各包含25个数值的两个组距？

不论你处理的数据集中数值的规模有多大，建立组距都要参照如下一些一般原则。

1.选择一个包含2、5、10或20个数据点的组距。在我们的案例中，选择了包含5个数据点的组距。

2.选择一个组距，使10~20个这样的组距就可以覆盖所有的数据。完成这一步的简便方式是计算极差，接着除以你想使用的组距的个数（10~20）。在我们的案例中有45（47-2）个数值，而且我们想用10个组距：45÷10≈5，这就是每一个组距的规模。如果你有一个值域范围是100~400的数据集，想用20个组距：300/20=15，因此15就是组距。

3.根据组距把所有分组罗列出来。在之前给出的频数分布中，组距是5，而且我们的最低组距以0开始。

注意，最大的组距处在频数分布的顶端。

一旦建立了组距，就着手完成频数分布的频数部分。这就是简单地计算原始数据中每一个数值出现的次数，接着将这个数字放入这个数值代表的组距。

在我们之前建立的频数分布中，在30和34之间出现的数值的个数或在30~34的组距中数值的个数是8。因此，8写入频数所对应的列。这就是频数分布。



先图示数据然后再进行所谓的计算或分析有时候是一个好主意。通过“看”数据，你可能会认识变量之间的关系，知道什么样的描述统计值是你描述数据正好要使用的，等等。接下来的工作能提高你对所做工作的洞察力和价值。

图形密度：建立直方图

现在，我们已经得到了多少数值落入对应组距的记录，接着就进入下一步并建立**直方图**（histogram）——频数分布的形象表示，图中频数以条形表示。



依据你所读的书和使用的软件，数据的形象表示就是图（graph或chart）。实际上两者没有差别。你需要了解的就是图是数据的形象表示。



依据下面的步骤建立直方图。

1—使用制图纸，在x轴等距离地列出数值，如图4.1所示。现在确定组距的组中值，也就是位于组距中间的数据点。组中值很容易计算，一眼就看得出来，但是你也可以简单地将组距的最大值和最小值加起来然后除以2。例如0~4组距的组中值是0和4的平均数，也就是4/2=2。

2—围绕每一个代表组距的组中值绘制以高度代表这个组距频数的条形或柱形。例如在图4.2中，你可以看到的第一个条形，也就是0~4组距是由频数1代表（表示1乘以0到4之间的一个值）。继续绘制条形或柱形直到每一个组距的频数都用图形表示。

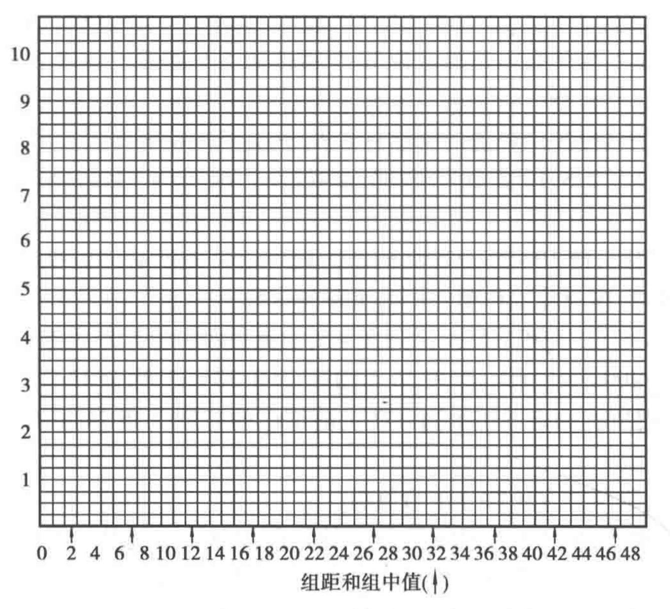


图 4.1　沿 x 轴的组距

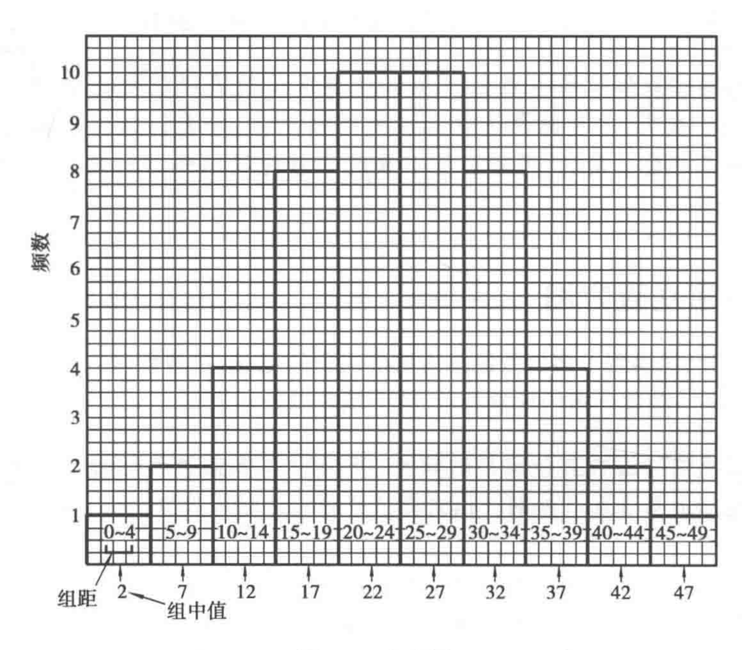


图 4.2　直方图

图 4.2 是一幅很好的直方图，表示我们之前使用的 50 个数值的频数分布。注意每一个组距是如何由 x 轴的一定值域代表的，也是组距的规模，在这个例子中组距规模是 5。

点算计数方法

通过上页的简单频数分布图就可以知道，这比起简单地列出数据，能让你对数值分布了解更多。你会对有什么数据和数据发生的频数有很好的了解。

但还有另外的图形可以展示每一个数值发生的次数,如图 4.3 所示。我们使用点算特定组距所包含的数值的频数的方法。这种方式可以更直观地表示一个数值相对于另一个数值出现的次数的多少。

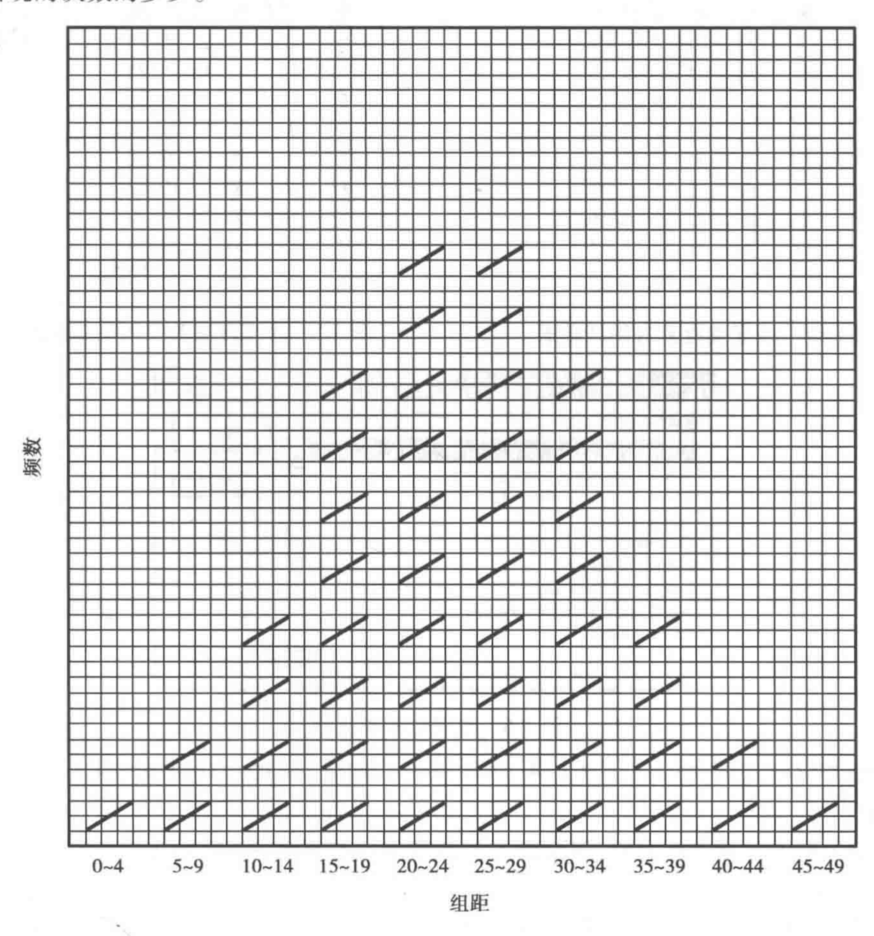


图 4.3 点算数值

使用分析工具库建立直方图

在 Excel 中使用分析工具库的直方图工具建立直方图。需要完成三步:

1.确定单元格,使用单元格包含的数据建立直方图。

2.确定“数据间隔”或者范围(也就是我们定义的组距),然后输入数据。

3.从工具库中选择直方图工具。

你已经上道了,现在试着自己来完成。先从在新的工作表录入数据开始(见图 4.4)。

1.输入你要使用的数据间隔(见图 4.5)。你不需要输入数据间隔(或者一个组距)的范围,只需要输入数据间隔的起始数据(例如 4),接着下一个数据间隔就是下一个组距的起始数据(例如 9)。你可以看到单元格 G2 到 G11 的数据间隔,对应的列标签我们已经输入了。

2.在工作表中选中空白单元格,直方图将放在这里。

3.单击[数据]→[数据分析]，如图 4.6 所示的数据分析对话框。

	A	B	C	D	E
1	Scores				
2	47	10	31	25	20
3	2	11	31	25	21
4	44	14	15	26	21
5	41	14	16	26	21
6	7	30	17	27	24
7	6	30	16	29	24
8	35	32	15	29	23
9	38	33	19	28	20
10	35	34	18	29	21
11	36	32	16	27	20

图 4.4　用于建立直方图的数据

	A	B	C	D	E	F	G
1	Scores						Bins
2	47	10	31	25	20		49
3	2	11	31	25	21		44
4	44	14	15	26	21		39
5	41	14	16	26	21		34
6	7	30	17	27	24		29
7	6	30	16	29	24		24
8	35	32	15	29	23		19
9	38	33	19	28	20		14
10	35	34	18	29	21		9
11	36	32	16	27	20		4

图 4.5　为直方图的应用建立数据间隔

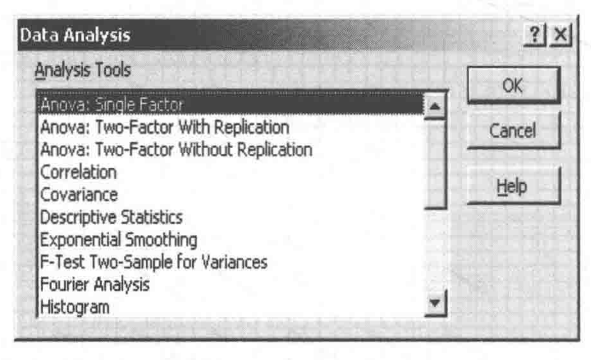


图 4.6　数据分析对话框

4.向下滑动直到看到直方图工具，然后双击。这样，你就会看到如图 4.7 所示的直方图对话框。

5.单击输入区域选框，拖动鼠标选定包含数据的单元格范围，在我们的例子中是单元格 A2 到 E11。

6.单击接收区域选框，拖动鼠标选定包含数据间隔的单元格范围，在我们的例子中是单元格 G2 到 G11。

7.单击输出区域选项。输入放置直方图的单元格位置。一个单元格就可以。在我们的例子中，我们输入单元格 H1。完成的建立直方图对话框，如图 4.8 所示。

8.单击[累计百分率](Cumulative Percentage)和[图表输出](Chart Output)选项框。第一个选项给出累计百分比，第二个选项给出图表。

9.单击确定，然后就得到图 4.9 所示结果：一幅简洁的直方图。

使用图 4.9 所示的直方图有以下 3 项注意事项。

第一，Excel 建立数据间隔始终从最小的数据间隔数值开始。

第二，Excel 建立了两列数据，一列的标签是“数据间隔”；另一列的标签是“频率”。

第三，Excel 建立了一个标签为“其他”的分类项，任何不在数据间隔包含范围内的数值都会在这个分类项中出现。例如，如果在数据集中有一个超出间隔范围的数值(比如 87)，就会在“其他”中出现数值 1(如果没有超过数据间隔范围的数值，“其他”中的数值就是 0)。

图 4.7　直方图对话框

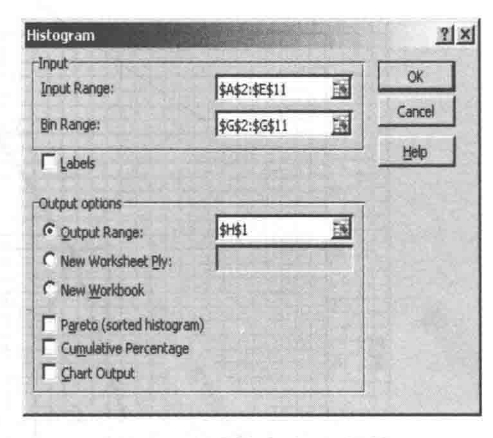
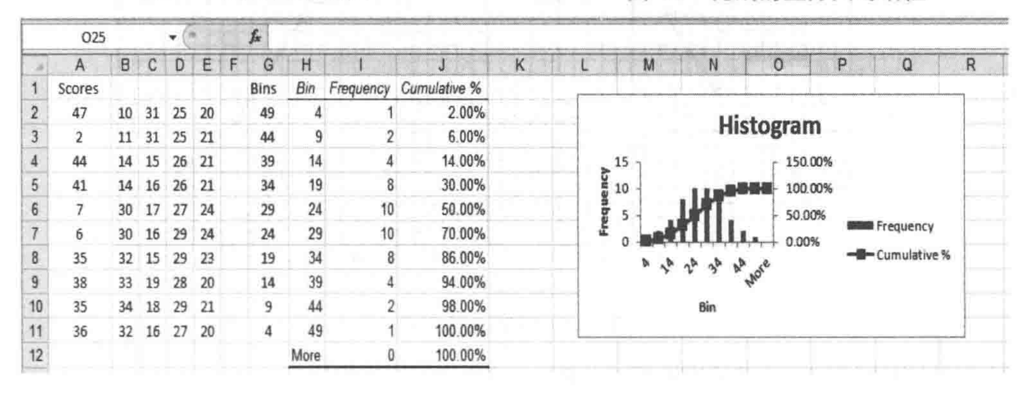


图 4.8　完成的直方图对话框



	A	B	C	D	E	F	G	H	I	J
1	Scores						Bins	Bin	Frequency	Cumulative %
2	47	10	31	25	20		49	4	1	2.00%
3	2	11	31	25	21		44	9	2	6.00%
4	44	14	15	26	21		39	14	4	14.00%
5	41	14	16	26	21		34	19	8	30.00%
6	7	30	17	27	24		29	24	10	50.00%
7	6	30	16	29	24		24	29	10	70.00%
8	35	32	15	29	23		19	34	8	86.00%
9	38	33	19	28	20		14	39	4	94.00%
10	35	34	18	29	21		9	44	2	98.00%
11	36	32	16	27	20		4	49	1	100.00%
12								More	0	100.00%

图 4.9　直方图工具输出结果

下一步：频数多边形图

建立直方图或点算图不困难，而下一步（下一个用图形说明数据的方式）更容易。我们准备使用相同的数据——实际上就是刚才建立的直方图 4.2——建立频数多边形图。**频数多边形图**（frequency polygon）是代表组距内数值频数的连续线段，如图 4.10 所示。



如何绘制频数多边形图？按照下面的步骤来做。

1—在直方图的条形或柱形的顶端设置组中值。（见图 4.2）

2—用线段连接组中值就得到频数多边形图。

为什么使用频数多边形图而不是直方图来表示数据？因为频数多边形图是更优的选择。频数多边形图看起来比直方图（表示频数变化的直线看起来总那么规整）更生动，但基本上涵盖的是相同的信息。

累计频数

如果建立了频数分布，而且已经用直方图或频数多边形图对那些数据进行了形象的展示，那么还有一种选择就是建立组距内数值的累计发生频数的形象图示，即**累计频数分**

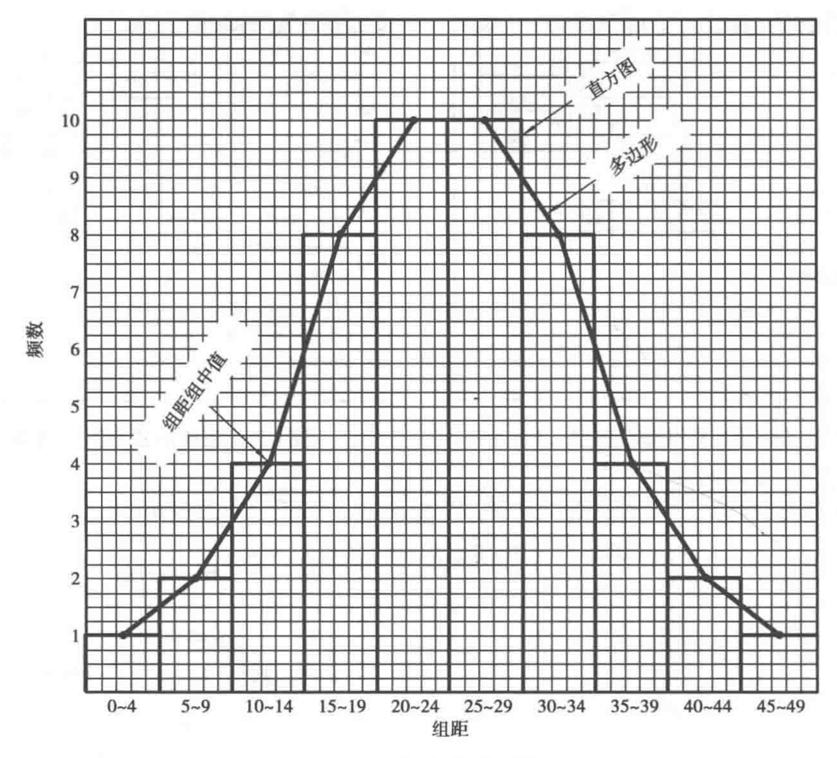


图 4.10 频数多边形图

布(cumulative frequency distribution)。

累计频数分布的基础就是数据的频数分布，但是多增加了一列（累计频数），如下表。

组 距	频 数	累计频数
45~49	1	50
40~44	2	49
35~39	4	47
30~34	8	43
25~29	10	35
20~24	10	25
15~19	8	15
10~14	4	7
5~9	2	3
0~4	1	1

先建立标签为累计频数的新的一列。接着将一个组距的频数和其下的所有频数加起来。例如，0~4 组距的发生频数是 1，在这一组之下没有其他组距，因此累计频数是 1。对于 5~9 组距来说，这一组的发生频数是 2，这一组之下的发生频数是 1，因此这一组和这一组之下的总的频数是 3(1+2)。最后的组距(45~49)发生频数是 1，这一组或这一组之下的总的频数是 50。

一旦我们建立了累计频数分布，那么数据就可以表示为直方图或频数多边形图。这一次，我们直接以每个组距的累计频数对应组距中点绘制。你可以在图 4.11 看到以这一章开始提供的 50 个数值为基础建立的累计频数分布。

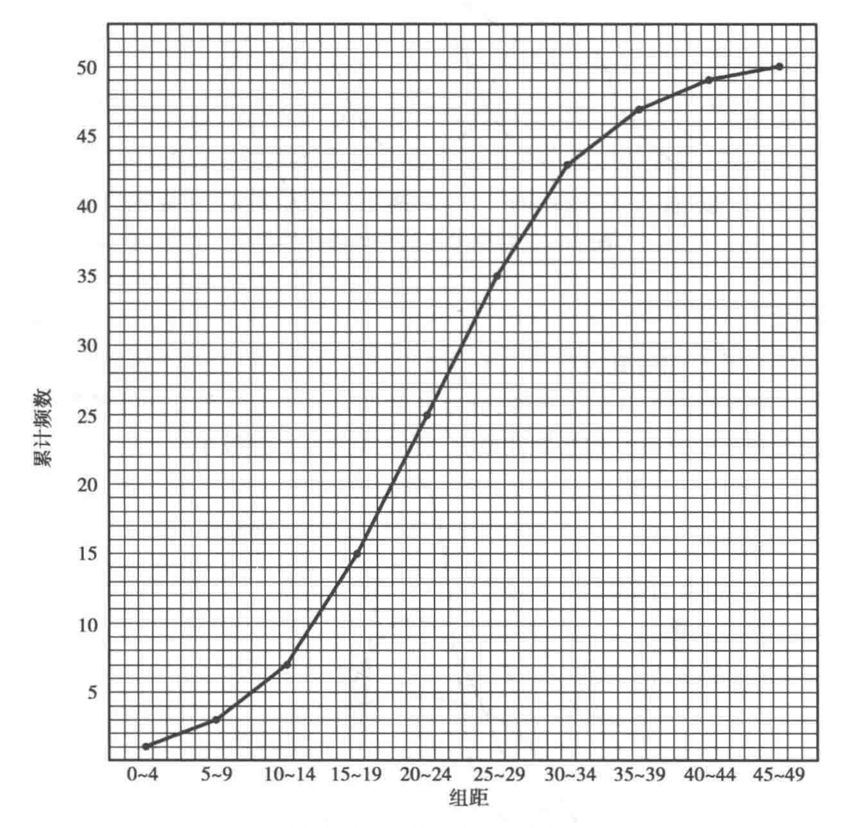


图 4.11 累计频数分布图



累计频数多边形图的另一个名称是**肩形图**(ogive)。如果数据是正态分布(相关内容见第 8 章),肩形图呈现出的就是更为大家所熟悉的钟形曲线。

扁平和狭长的频数分布

现在你可以一定程度地推断出数据分布相互之间在许多方面非常不同。实际上,有四个方面不同:平均值、变异性、偏度*和峰度。后面两项是新的概念,我们会在图形展示这两项时再进行定义。现在依次定义这四项特征值,并且用图形说明。

平均值

我们再次回顾集中趋势量数。如图 4.12 所示,你会看到三个不同的数据分布的平均值的差异程度。你会看到数据分布 C 的平均值大于分布 B 的平均值,分布 B 的平均值大于分布 A 的平均值。

* 也叫倾斜度。——译者注

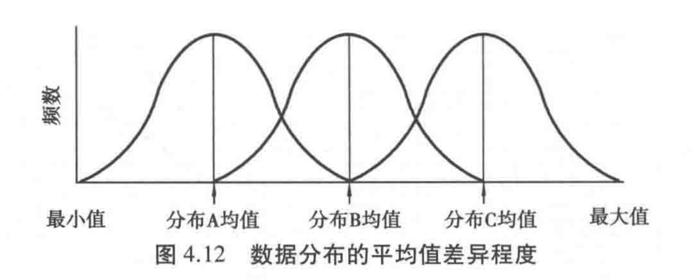


图 4.12 数据分布的平均值差异程度

变异性

在图 4.13 中你可以看到具有相同平均值但是变异性不同的三个数据分布。数据分布 A 的变异性小于数据分布 B 的变异性，分布 B 的变异性小于分布 C 的变异性。换句话说，就是三个数据分布中分布 C 的变异性最大，分布 A 的变异性最小。

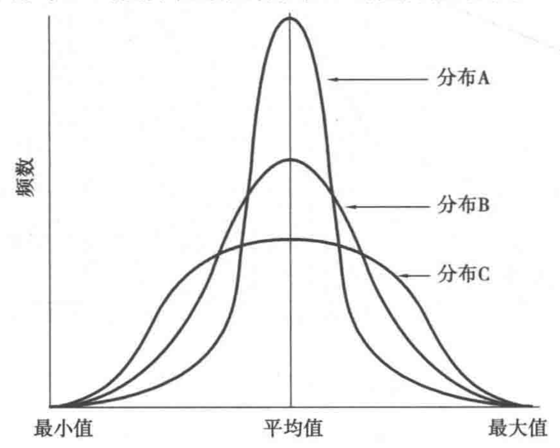


图 4.13 数据分布的变异性差异程度

偏 度

偏度(skewness)是对数据分布对称性缺失或者分布不平衡的测量。换句话说就是分布的一个"尾巴"比另一个长。例如在图 4.14 中，数据分布 A 的右侧尾比左侧尾长，相应的就是分布的大数值一端的发生次数较小。这是正偏度分布。以下案例就是这种情况：一次难度非常大的考试，只有少数学生取得相当高的成绩，而大多数学生的成绩相当低。分布 C 的右侧尾比左侧尾短，相应的就是分布的大数值一端的发生次数较大。这是负偏度分布，难度较小的考试就是这种情况(大多数是高分，少数是低分)。分布 B 刚好左右尾相等，没有偏度。如果均值大于中位数，就是正偏度分布。如果中位数大于均值，就是负偏度分布。

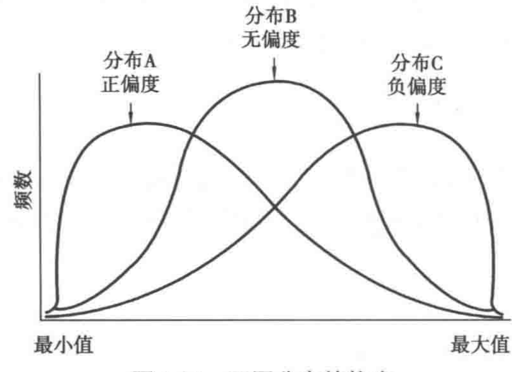


图 4.14 不同分布的偏度

峰 度

虽然这个词听起来像个医学条件,但确实是我们区分数据分布的相互区别程度的最后一个指标。**峰度**(kurtosis)与数据分布看起来是扁平还是陡峭有关,而且用于描述这个特征的概念是相对的。例如,**低阔峰**(platykurtic)一词是指数据分布相对于正态分布或钟形分布来说扁平。**高狭峰**(leptokurtic)一词是指数据分布相对于正态分布或钟形分布来说陡峭。在图 4.15 中,数据分布 A 与分布 B 相比是扁平的。数据分布 C 与分布 B 相比是陡峭的。图 4.15 看起来与图 4.13 类似,这是由于低阔峰数据分布比不是低阔峰的数据分布更分散。类似地,陡峭峰数据分布相对于其他分布来说离散性或变异性更小。

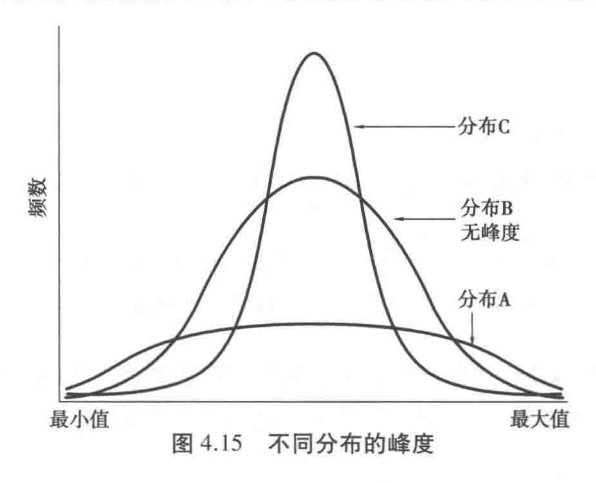


图 4.15 不同分布的峰度

> **更多 Excel**
>
> Excel 可以帮助计算一组数据的偏度值和峰度值。你可以用 SKEW 函数与 KURT 函数计算。例如,如果你在单元格 A1 到 A20 有 20 个数值,函数形式如下:
>
> =SKEW(A1:A20)
>
> = KURT(A1:A20)
>
> 只要在你选定的呈现结果的单元格输入公式即可。

Tech Talk 偏度和峰度是常用的描述用语(例如,"这是负偏度分布"),有一些数学指标可以测定分布的倾斜或陡峭的程度(实际上 SKEW 函数和 KURT 函数计算的就是这些分布特征的测量值)。偏度是通过均值减去中位数计算。如果数据分布的均值是 100 而中位数是 95,偏度值就是 100-95=5,这个数据分布是正偏度分布。

如果数据分布的均值是 85 而中位数是 90,偏度值是 85-90=-5,这个数据分布是负偏度分布。其实还有更复杂的公式,这个公式因为考虑了数据分布的标准差,所以计算出的偏度值能够相互比较。

$$sk = \frac{3(\overline{X} - M)}{s} \tag{4.1}$$

其中：*sk* 是皮尔逊的（他也计算相关系数，你会在第 5 章了解到）偏度量数，

$\overline{X}$ 是均值，

M 是中位数。

使用这个公式可以在绝对意义而不是相对意义上比较不同数据分布的偏度。例如数据分布 A 的均值是 100，中位数是 105，标准差是 10。数据分布 B 的均值是 120，中位数是 116，标准差是 10。使用皮尔逊的公式，分布 A 的偏度是-1.5，分布 B 的偏度是 1.2。分布 A 是负偏度，分布 B 是正偏度。不管偏度的方向如何，分布 A 比分布 B 更倾斜。注意事项：公式 4.1 不是 Excel 使用的公式，我们介绍这个公式是因为它比 Excel 使用的公式更容易理解。

最好的图表

接下来内容会非常有趣。

本章这部分内容中，我们主要展示如何使用 Excel 在短时间（实际上应该按秒计，仅仅是三秒，或者说三步）之内建立一个简单的图表。而且我们也会向你展示，如何让这个图表好看，并且包含更多的信息。我们将建立图 4.16 所示的柱形图，表示的是参与调查的男性和女性的数量。接着向你展示如何进行一些简单但是高效的编辑。



为什么图 4.16 中的图叫作柱形图而不是条形图？Excel 设计人员将条形是垂直的图命名为柱形图，将条形是水平的图命名为条形图。很简单，就是这样。

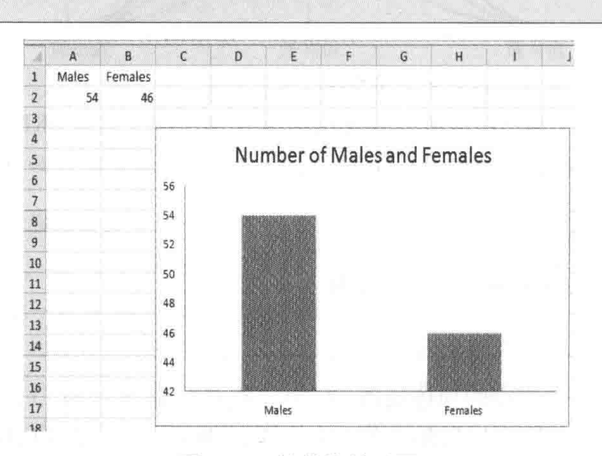


图 4.16　简单的柱形图

建立 Excel 图表需要记住以下几点：

1.Excel 会使用你选中的数据，但是你必须首先输入你想用于建立图表的数据。例如，你有一列数据，想用平均数建立图表，那么你需要计算平均数，然后再建立图表。

2.如图 4.16 所示，Excel 用工作表中的数据建立了图表，当你保存包含图表的工作表时，你既保存了用于建立图表的数据，也保存了这个图表。

3.当用于建立图表的数据发生变化时，图表本身也会变化。这两个东西（数据和图表）是息息相关的。

4.图表可以很容易被剪切并粘贴到其他地方(稍后会展示)。

5.Excel 提供了非常多不同的图表形式,所有的这些图表都可以通过单击几次鼠标来建立,而且可以按照你看着合适的标准来调整。

在开始轻松地建立图表之前还需要记住一点:建立图表容易,而建立好的图表不容易。图表必须承载我们从这章开始就在讨论的所有内容。因此打起精神来,按照这些规则和原理来建立图表。但是也不要为了图表而制作图表,忘记了图表代表着什么内容。

你的第一个 Excel 图表:需要铭记的时刻(舒口气吧)



按照下面的步骤建立 Excel 图表。

1.选中你用于建立图表的数据。在图 4.17 中,我们将单元格 A1 到 B2 高亮显示。

2.单击[插入](Insert),在选项中选择你要使用的图表类型。在图 4.17 中,我们选择了[柱形图](Column)。

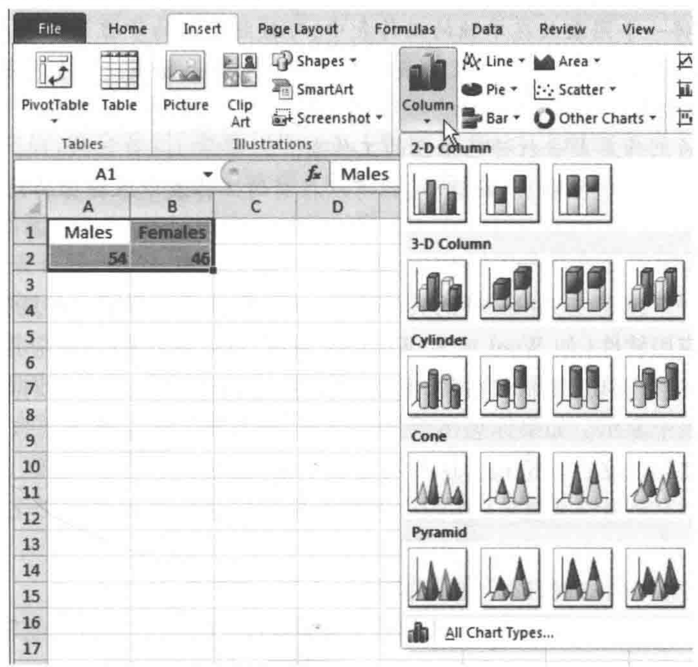


图 4.17　选择要建立的特定图表类型

3.单击[二维柱形图](2-D Column)选项下的第一个按钮,马上就得到如图 4.16 所示的柱形图了。只需单击几下,一个非常漂亮、整齐的柱形图就会出现。真是太简单了!

更多 Excel

在工作表中选定图表后，图表四周就会出现操作点，只要拖动这些操作点就可以很容易地改变图表的大小。如果你拖动位于图表上部或者下部的操作点，就会改变图表的高度；如果你拖动位于图表左侧或者右侧的操作点，就会改变图表的宽度。这两种情况下，你只能改变一个维度，图表就会变得有点怪。如果你拖动位于顶角上的操作点，两个维度就会同时发生变化。如果你想改变图表在工作表中的位置，只要将光标放在图表上，当光标指针变成了十字准星，就可以将图表拖到你想放的位置。

喏，建立图表非常容易，只要单击三次就可以完成。如图 4.17 所示，我们可以选择选项中任何一种类型的图表（或者选择其他的图表选项）。当然，一个图表应该适合其数据类型，但是 Excel 中肯定有你需要的图表类型。

更多 Excel

如果你想将一个图表放在单独的工作表中，可按照下面的步骤来完成：

1.在图表区单击右键（不是图表的某一要素如任何一条轴或者线）选中图表，接着在菜单中选择[复制]。

2.建立新的工作表或者打开已建立的工作表。

3.单击右键，单击[粘贴]。就这样，你可以在新的工作表中依据你的意愿移动图表。

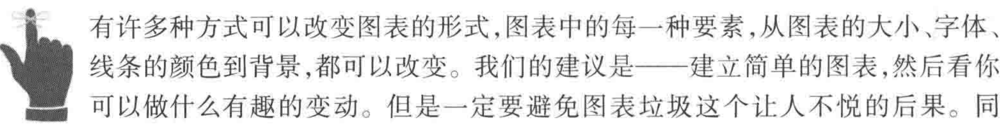
图表一旦建立，就和 Excel 的其他要素一样，很容易被剪切或者复制，然后粘贴到新的应用软件（如 Word 或者 Windows 系统、Mac 系统的其他应用软件）中。但是，如果你直接进行剪切/复制和粘贴，当建立图表的数据发生变化时，数据变化不会在这些被粘贴的图表中显示。如果你想体现这种变化，就需要将建立图表的数据和不论在哪里出现的图表之间建立自动链接。这种功能很不错，但是要小心，一旦建立自动链接，不论这个图表被粘贴到什么位置，工作表中的数据变化都会导致图表的变化。按照如下步骤来完成：

1.选中图表。单击右键，选择复制选项。

2.转换到你想粘贴图表的 Word 文档或者 Microsoft Office 的任何其他软件（以及其他非微软的软件），如 PowerPoint。

3.单击右键出现选项卡，从粘贴选项中选择[保留源格式]选项。应用这种方式，不论在 Excel 中建立的图表发生什么变化，在新的应用软件中的图表都会发生相应的变化。

好图表的第二步：让图表更美观



有许多种方式可以改变图表的形式，图表中的每一种要素，从图表的大小、字体、线条的颜色到背景，都可以改变。我们的建议是——建立简单的图表，然后看你可以做什么有趣的变动。但是一定要避免图表垃圾这个让人不悦的后果。同时，始终要保存你的原始图表（使用别的名称来保存你修改的图表），这样你可以在需要的

时候回到原始图表(如果你进行的改变太多,我们相信你肯定会需要回到原始图表)。

如果你想在图表中突出一个特定的方面,色彩是最好的选择。但是如果你不能恰当地使用色彩,那就可能是花里胡哨。有时,黑白打印机打出来的色彩非常奇怪,而彩色打印非常贵。这时色彩使用就变得不必要了。因此,始终保持图表简单,仅仅在需要的时候使用色彩,并且确保你可以很好地表现色彩。



按照这些步骤,调整图4.18所见的图表形式。改变一个图表中任何一个要素的表现形式的一般规则是:双击这个要素,接着进行任何你需要的调整。你也可以使用图表工具(Chart Tool)(你马上会看到)来调整,或者单击右键然后选择相应的选项。

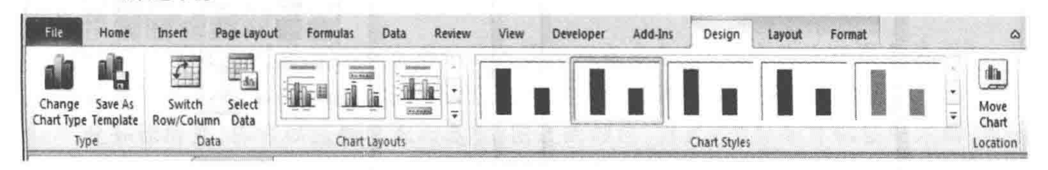


图4.18 图表工具

记住,一旦在图表上单击一下,进行修改的图表工具就会出现,如图4.18所示。图表工具菜单出现之后,就会看到[类型](Type)、[数据](Data)、[图表布局](Layouts)、[图表样式](Styles)、[位置](Location)选项。我们的重点是图表布局和样式。

1.单击图标选中图表,这样图表工具就出现了。

2.选择[图表布局]选项,单击任何一个布局(这个例子是单击图表标题)。接着在图表标题上单击,就可以看到如图4.19所示的内容,Excel可以让我们使用不同的工具编辑图表的标题,包括字体和颜色等。

3.输入你想要的标题,再按你的要求编辑。

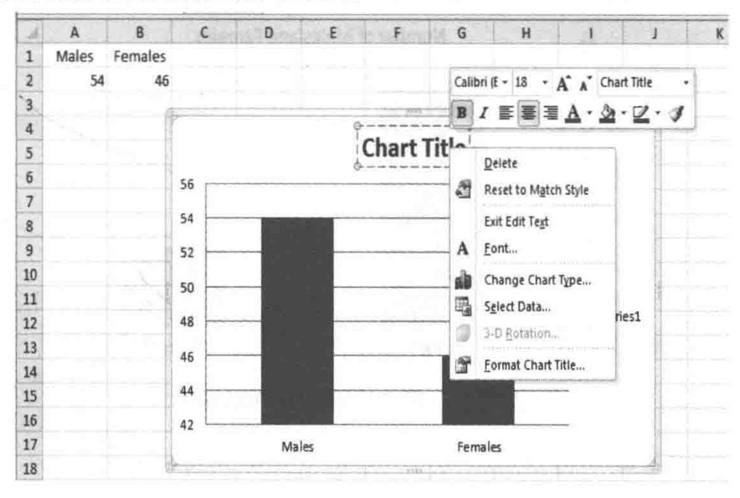
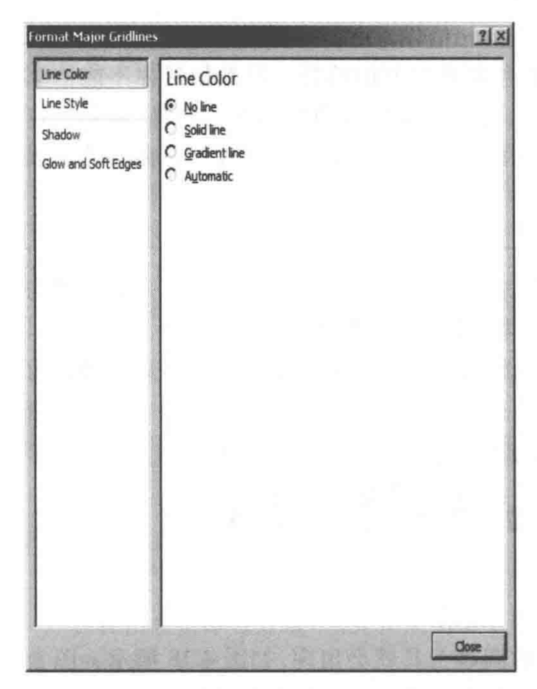


图4.19 从图表布局选项中选择图表标题项

4.接着来取消水平网格线,在图表中的任何一条水平线上右键单击,选择设置网格线格式选项,然后选择无线条选项,见图4.20所示的设置主要网格线格式对话框。



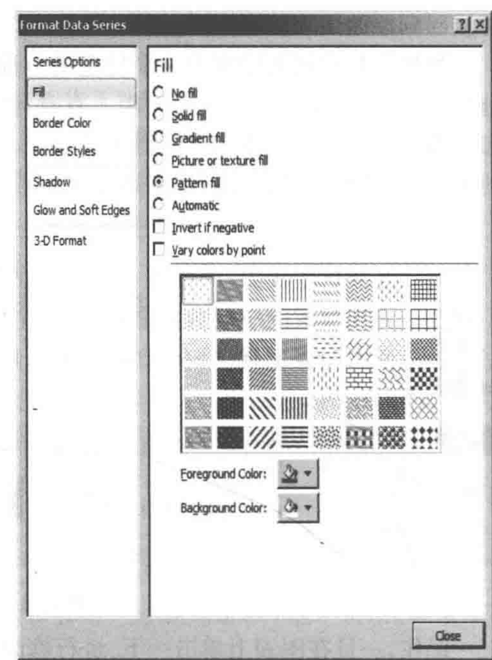


图 4.20　取消图表的网格线　　　图 4.21　改变图表每个要素的类型

5.最后，我们可以改变表示数据的条形的类型。如果要进行改变，在你要调整的柱形上单击右键，选择设置数据系列格式，在设置数据系列格式对话框中选择填充选项，然后单击图案填充项，如图 4.21 所示。

6.最后，我们单击图例 Series 1（系列一）并选择删除，因为图表中只有一个要素（"发生次数"），所以不需要图例。你可以在图 4.22 看到最终的图表。

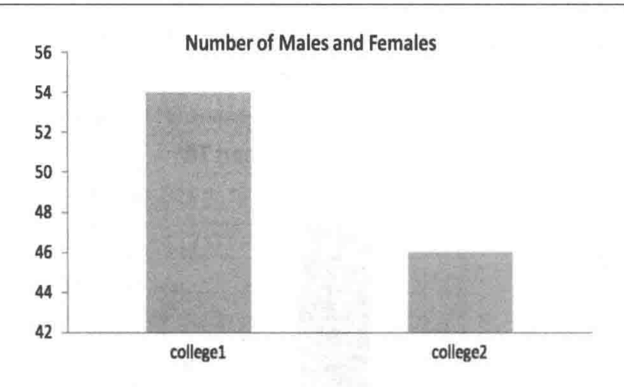


图 4.22　简单的图表，简单的改变

更多 Excel

下面的内容你也许认为非常有用,也许觉得多余——用图像做图表的元素。例如,在一个条形图中不用条块来表示班上男生、女生的数量,而是用其简绘人像表示。这样会生动、直观得多。按照下面的步骤可以在图表中使用图像。

1.建立图表。

2.单击你要用图像替代的图表要素。

3.在选定的要素上单击右键,然后选择[设置数据系列格式]选项。

4.在设置数据系列格式对话框的填充选项中,单击图片或纹理填充选项。

5.单击[文件]按钮,找到你需要的文件。你应该已经准备了要使用的图像,如果你的电脑中还没有,就下载。

6.单击[关闭],你就可以看到图像替代了柱形或者条形(或者其他的图表类型)。你可能需要做一些工作来让这些图更简洁,这是非常好的说明数据结果的方式。

就如你所见,你可以选择填充的类型,图表的颜色、样式、阴影等可调整的方面,或者更换标题、坐标轴等,你可能会花费很长的时间来调整这些。但重点是要使得图表更清晰。我们建议:试试所有选项,这样的话我们可以充分了解这些选项的功能,以及什么时候该应用它们。

其他用图表显示数据的绝妙方法

本章到现在为止我们所做的就是给出一些数据,然后说明如何使用图表(如直方图和多边形图)来形象地展示数据。但是行为和社会科学领域还使用一些其他类型的图表,虽然对你来说没必要确切地知道如何建立或绘制这些图表,但是你至少应该熟悉它们的名称和应用。因此下面介绍这些流行的图表,以及如何应用和实现。

条形图

条形图和柱形图一样,但是在条形图中,分类项在垂直轴 y 轴上,数值在水平轴 x 轴上。

线　图

当你想用相同的间距表示数据的趋势,就用线图。下面的例子你可能会使用线图:

- 三个州立大学的大学生中每季出现单核白血球增多症病案的数量
- 每学年入学人数的变化
- 两条航线每季的旅行人数

图 4.23 是三个州立大学的大学生中每季汇报的单核白血球增多症病案的数量的线图。

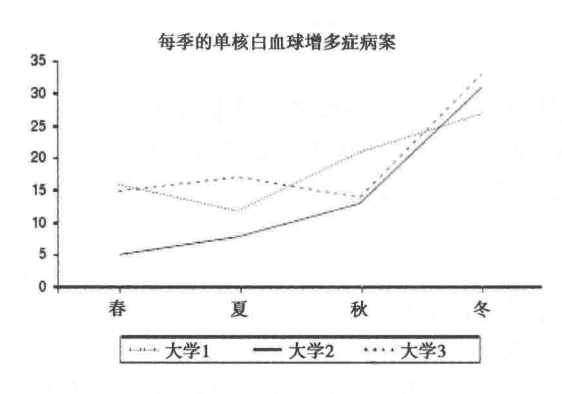


图 4.23　使用线图说明随时间变化的趋势

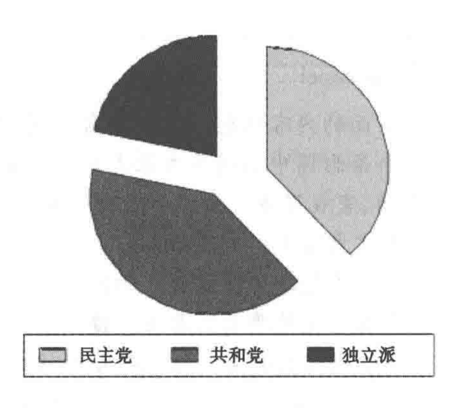


图 4.24　使用饼图显示不同分类项的比例

饼　图

如果你想表示一系列数据点组成的项目的比例，可以使用饼图。下面的例子你可能想要使用饼图：

- 生活贫困的儿童中各种族的比例
- 注册夜间部和日间部的学生的比例
- 不同性别的参与者的年龄分布

在图 4.24 中，你可以看到不同政党的投票人数的饼图。我们把饼图分成三块（只是为了好玩儿），不过这个怎么看都挺不错。

更多 Excel

在 Excel 中建立图表时，你必须知道何时单击左键、何时单击右键。例如制作饼图，我们应该单击右键选择整个图表来改变填充模式，但是应单击左键来选择图表的某个要素。

这是核心内容

现在应该开始学习新内容（最起码对于学习统计学的新手来说是新的）——数据透视表。

数据透视表可以让你在不使用任何公式和函数的情况下，在数据表或数据列中看到汇总信息。数据透视表（pivot table）有一个词——pivot（枢纽），表示你要重新安排或者移动工作表的行和列中包含的数据（也要组织 Excel 中的许多工作表）。

建立数据透视表听起来很吓人，这也是人们不太愿意学习和使用的原因之一。但是，如果我们从简单的开始，先建立一个有用的分析表，并展示一些基本的形式——你可以以此起步，拓展自己建立和使用更复杂的数据表的技能。

这是一个非常简单的案例。我们有一个 20 人、每人两个数据点的数据集，如图 4.25 所示。第一个变量是性别（Gender）（1＝男性，2＝女性），第二个是他们的学习项目类

	A	B	C
1	ID	Gender	Category
2	1	1	A
3	2	2	B
4	3	2	C
5	4	2	C
6	5	1	B
7	6	1	B
8	7	1	A
9	8	2	A
10	9	2	B
11	10	1	B
12	11	1	B
13	12	2	B
14	13	2	A
15	14	1	A
16	15	2	A
17	16	2	A
18	17	1	A
19	18	1	B
20	19	2	C
21	20	2	C

图 4.25　数据透视表的数据

型(Category)(A、B 和 C 项目)。

按照下面的步骤建立数据透视表。

1.选择你想用于这个分析的数据。在这里的例子中,我们选择了单元格 B1 到 C21 的数据,我不想使用 ID(编号)这列的数据。

2.在[插入]菜单,选择在屏幕左侧的[数据透视表]选项。接着在下拉菜单单击[数据透视表]项,就会看到如图 4.26 所示的建立数据透视表对话框。

3.选择表格出现的位置。在这个例子中,我们将表格放在数据所在的工作表(单元格 F2)。

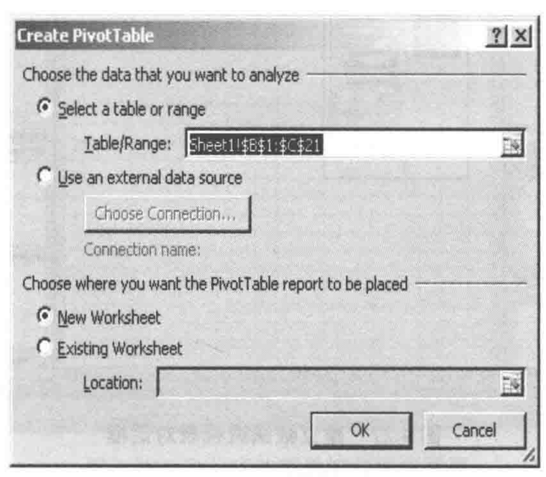


图 4.26 建立数据透视表对话框

4.单击[确定],你就会看到图 4.27 所示的分析表的开始,以及数据透视表字段列表(见右侧)。你可以看到从数据中选择的变量是 Gender 和 Category。

5.在数据透视表字段列表中将变量 Gender(性别)拖到行标签框中。完成之后,你就会看到在工作表的主要区域会形成一个数据表,如图 4.28 所示,性别(数据表的行)出现的同时,变量的两个变值 1 和 2 也出现。

6.现在,将 Category 拖到列标签框中。

7.最后,将 Category 拖到数据透视表字段列表的右下侧的[数值](Values)框中。啃啃,你可以看到如图 4.29 所示的数据透视表了。

不是很难,对吧?我们现在得到的哪些内容是之前没有的?只要单击几下鼠标,我们现在就可以知道如下的信息。

- 不同性别的人数(类别 1 是 9 人,类别 2 是 11 人)。
- 不同学习项目的人数(A 项目是 8 人,B 项目是 8 人,C 项目是 4 人)。
- 性别和学习项目的交互分类项的人数(性别 1 参加项目 B 的人数是 5 人)。

数据透视表的总体思想是使用现有的数据建立一个数据表,可以准确地展现你感兴趣的信息。一旦你学会使用数据透视表,你会非常惊奇地发现这些表是多么的有用。而且,当然了,一旦你将数据按照需要呈现在一个数据表中,只要单击几下鼠标,一个堆积柱形图就出现了。图 4.30 的堆积柱形图是依据图 4.29 所示数据建立的。

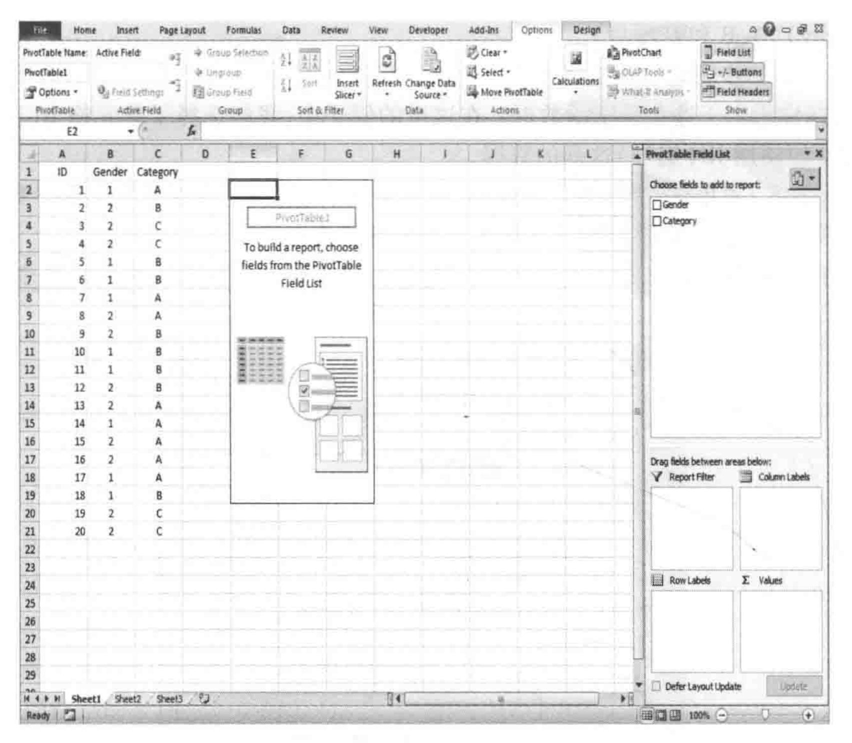


图 4.27　建立数据透视表对话框

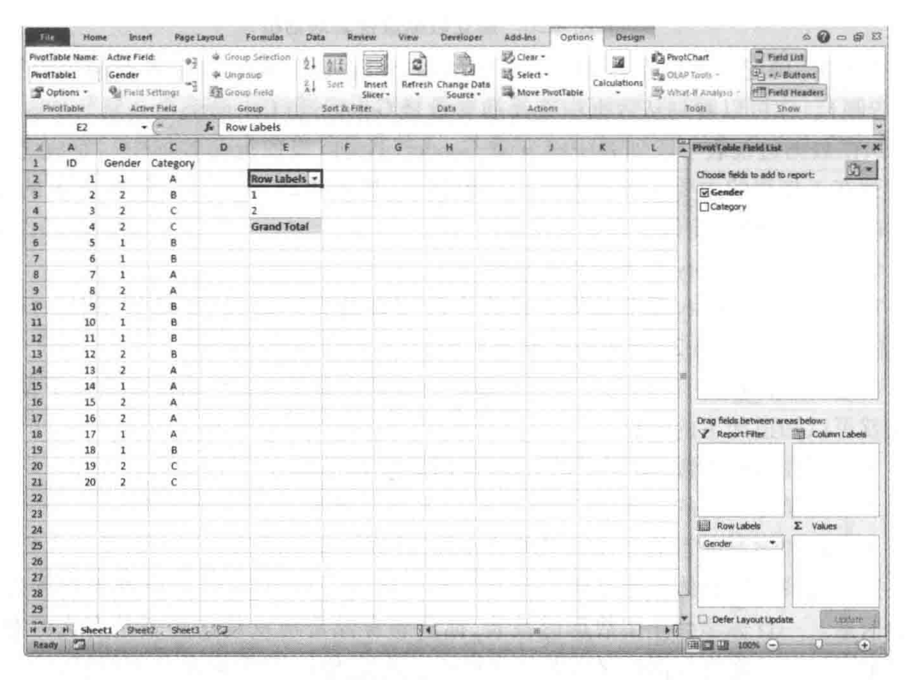


图 4.28　拖动变量建立数据透视表

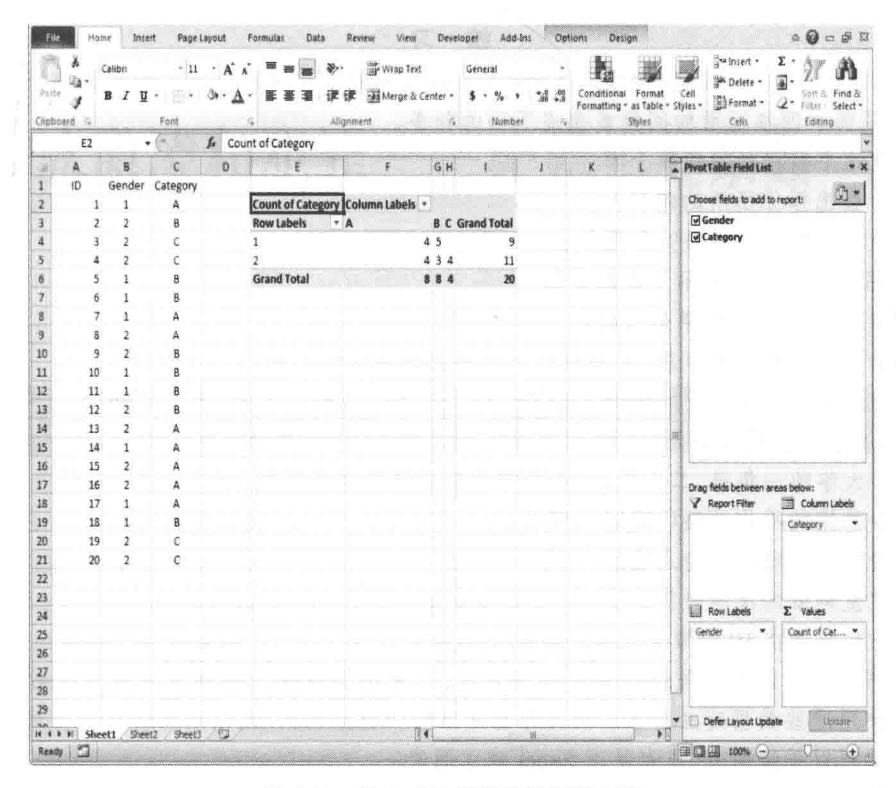


图 4.29 已完成的简单的数据透视表

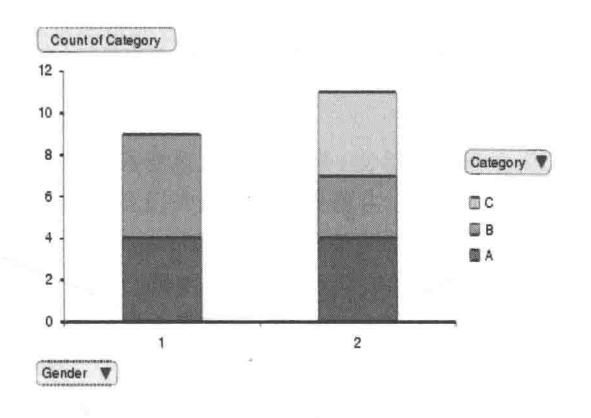


图 4.30 由数据透视表创建的堆积柱形图

小 结

毫无疑问建立图表很有趣，而且能极大地增加对数据的理解。依据我们在本章给出的建议可以很好地制作和使用图表，学会而且加强对数据的理解。

练 习

1.50 次阅读理解考试成绩(变量名 Comp Score)构成的第 4 章数据集 1 的数据集可以在附录 C 获得。回答下面的问题或完成下面的任务。

a.建立这个数据集的频数分布和直方图。

b.为什么选择你所用的组距?

c.这是有偏度的分布吗? 你是如何知道的?

2.你使用什么图表来展示下面的数据?

a.10 年间一个学区的学生人数。

b.参加课外活动的九年级、十年级和十一年级的学生的比例。

c.接种疫苗和没有接种疫苗的儿童的人数。

3.下面的数据,你使用饼图、线图、条形图还是柱形图表示,为什么?

a.某个大学的一年级、二年级、三年级和四年级学生的比例。

b.四个学期的平均成绩(GPA)的变化。

c.申请四个不同职位的人数。

d.对重复刺激的反应时间。

e.一个参与者在 10 个项目中每一项的不同得分。

4.去图书馆找一篇你感兴趣领域的期刊文章,文章包含调查数据但是没有对应数据的图表。使用这些数据建立图表。确定你要建立哪一类型的图表,以及为什么作出这样的选择。你可以手绘图表,或者使用 Excel 建立图表。

5.建立简单的条形图,说明冰激凌商店在春、夏、秋、冬四季的销售量。如果可以,使用冰激凌图像来代替条形。

6.建立你能够建立的看起来最坏的图表,如表格拥挤和文字无用。如果你的图表比你同学的都糟糕,那么你就赢了。

冰激凌和犯罪——计算相关系数 5

本章你会学到什么 ☺☺ ⊠⊠⊠⊠

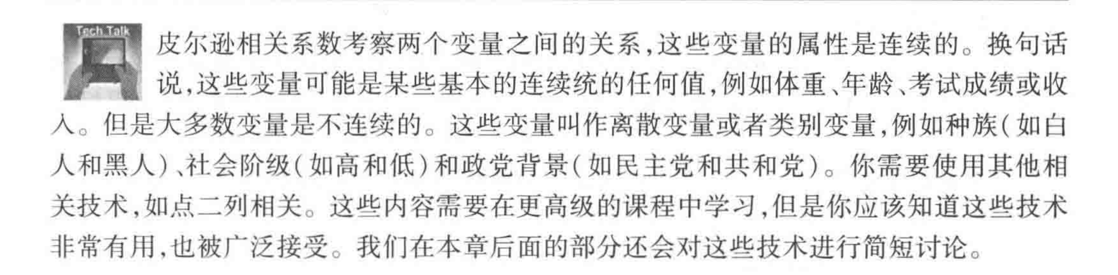
- 什么是相关系数,相关系数如何应用
- 如何计算简单的相关系数
- 使用 CORREL 函数计算相关系数
- 使用分析工具库中的 Correlation 工具计算相关系数
- 如何解释相关系数的值
- 其他类型的相关系数以及何时使用

相关系数到底是什么

集中趋势量数和变异性量数只是我们看好的用于描述数据集概况的两种描述统计量数。你已经了解,最具代表性的统计值(集中趋势)和散布或离散性(变异性)量数的值是描述数据分布特征的关键。

但是,我们有时感兴趣的是变量之间的关系,或者更准确地说,是当一个变量发生变化时,另一个变量如何变化。我们的这种兴趣表现在计算简单相关系数上。

相关系数(correlation coefficient)是反映两个变量之间关系的量化指标。这个描述统计值的值域范围是-1 到 1。两个变量的相关有时也叫作二元相关(bivariate correlation)。本章大部分讨论的相关类型是**皮尔逊积距相关**(Pearson product-moment correlation),是以它的发明者卡尔·皮尔逊命名的。



皮尔逊相关系数考察两个变量之间的关系,这些变量的属性是连续的。换句话说,这些变量可能是某些基本的连续统的任何值,例如体重、年龄、考试成绩或收入。但是大多数变量是不连续的。这些变量叫作离散变量或者类别变量,例如种族(如白人和黑人)、社会阶级(如高和低)和政党背景(如民主党和共和党)。你需要使用其他相关技术,如点二列相关。这些内容需要在更高级的课程中学习,但是你应该知道这些技术非常有用,也被广泛接受。我们在本章后面的部分还会对这些技术进行简短讨论。

还有其他类型的相关系数用于测量两个以上变量之间的关系,而且我们会把这些内容留到下一节统计课程。(你现在已经期待这样的课程了,对吧?)

相关系数的类型：选择 1 和选择 2

相关反映变量间关系的动态性质。这样做可以让我们理解变量发生变化时变化的方向是相同还是相反的。如果变量变化方向相同，相关是同向相关（direct correlation）或正相关（positive correlation）。如果变量变化方向相反，相关是反向相关（indirect correlation）或负相关（negative correlation）。表 5.1 给出了这些关系的汇总。

表 5.1 相关关系的类型和相应的变量之间的关系

变量 X 的变化	变量 Y 的变化	相关关系的类型	数 值	例 子
X 值增大	Y 值增大	同向的或正向的	正值，[0.00，1.00]	你用于学习的时间越多，考试成绩就会越高。
X 值减小	Y 值减小	同向的或正向的	正值，[0.00，1.00]	你在银行存的钱越少，所得利息就越少。
X 值增大	Y 值减小	反向的或负向的	负值，[-1.00，0.00]	你运动越多，体重就越轻。
X 值减小	Y 值增大	反向的或负向的	负值，[-1.00，0.00]	你完成考试的时间越少，所犯的错误越多。

要记住表中的例子反映的是通则。例如，考虑一次考试中完成试卷的时间和答对题目的数量的关系：一般来说，考试所用的时间越少得分越低。

得出这样的结论不是多难的事，因为回答得越快越可能有疏漏，如没有正确地阅读指导语。当然也有人既能答得快又能答得好，也有人虽然答得很慢但却不能回答得很好。重要的是，我们讨论的是一个群体在两个变量上的表现，是计算一个群体的两个变量之间的相关，而不是对应任何一个特定的个人。

需要记忆的内容

这里有许多与相关系数有关的简单但重要的内容。

相关系数的值域范围是[-1，1]。

相关系数的绝对值反映相关的强度。因此，相关系数-0.70比相关系数 0.50 表示的相关强度大。在考虑相关系数时，学生们常犯的一个错误是，只是因为正负号的原因而认为正相关总是比负相关强（例如“更好”）。相关适用于反映每个个体至少有两个数据点（或变量）的状况。

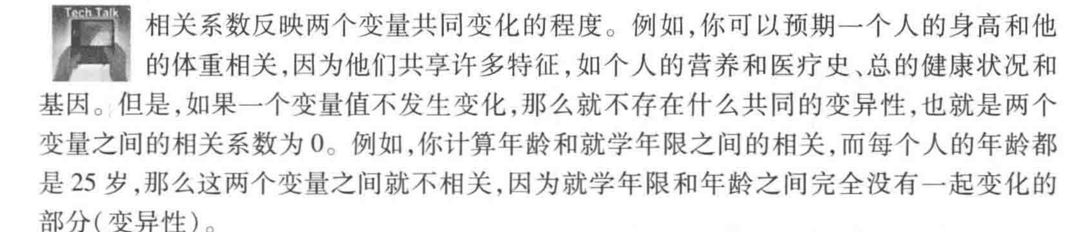
另一个易犯的错误是依据相关符号进行价值判断。许多学生认为负相关不好而正相关很好。如果他们仅仅从字面上理解，就会犯下“正相关”总是比“负相关”更强更好的错误。

皮尔逊积距相关系数用小写字母 r 表示，r 的下标表示相关的两个变量。例如：

r_{XY}是变量 X 和变量 Y 之间的相关系数。

$r_{weight.height}$是体重和身高之间的相关系数。

$r_{SAT.GPA}$是学术能力测验成绩（SAT）和平均成绩（GPA）之间的相关系数。



相关系数反映两个变量共同变化的程度。例如,你可以预期一个人的身高和他的体重相关,因为他们共享许多特征,如个人的营养和医疗史、总的健康状况和基因。但是,如果一个变量值不发生变化,那么就不存在什么共同的变异性,也就是两个变量之间的相关系数为 0。例如,你计算年龄和就学年限之间的相关,而每个人的年龄都是 25 岁,那么这两个变量之间就不相关,因为就学年限和年龄之间完全没有一起变化的部分(变异性)。

同样,如果限制或者控制一个变量的值域范围,这个变量和其他变量之间的相关系数(相对于变量的值没有限制的情况来说)会更小。例如你计算成绩优异学生的阅读理解成绩和年级之间的相关,你会发现相关系数小于你用全体学生的数据计算得到的同一相关系数。这是因为成绩优异学生的阅读理解成绩也非常好,相对于所有的学生来说这个成绩的变化很小。那么应该怎样避免这种情况?如果你对两个变量之间的关系感兴趣,就尽量收集充分的离散数据——这样你才可以得到最具代表性的结果。

计算简单相关系数

公式 5.1 所示是变量 X 与变量 Y 之间的简单皮尔逊积矩相关系数的计算公式:

$$r_{XY} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}} \tag{5.1}$$

其中:r_{XY}是 X 与 Y 之间的相关系数,

n 是样本规模,

X 是变量 X 的具体数值,

Y 是变量 Y 的具体数值,

XY 是每一个 X 值与相应的 Y 值的乘积,

X^2 是 X 值的平方,

Y^2是 Y 值的平方。

这是案例中使用的数据:

	X	Y	X^2	Y^2	XY
	2	3	4	9	6
	4	2	16	4	8
	5	6	25	36	30
	6	5	36	25	30
	4	3	16	9	12
	7	6	49	36	42
	8	5	64	25	40
	5	4	25	16	20
	6	4	36	16	24
	7	5	49	25	35
总计、合计或 $\sum$	54	43	320	201	247

在将数据代入公式之前，我们要确保你理解了每个符号代表的含义。

$\sum X$ 或 X 值的总和是 54。

$\sum Y$ 或 Y 值的总和是 43。

$\sum X^2$ 为每一个 X 值的平方的总和，是 320。

$\sum Y^2$ 为每一个 Y 值的平方的总和，是 201。

$\sum XY$ 为 X 和 Y 的乘积的总和，是 247。



一系列值的总和的平方与平方值的总和很容易混淆。一系列值的总和的平方，就如 2 和 3，先加起来（就是 5），接着将和值平方（是 25）。平方值的总和，就如 2 和 3，先平方（分别是 4 和 9），接着将平方值加起来（是 13）。当你比较时注意看括号内的数字。

按照下面的步骤计算相关系数：

1—列出每个参与者的两个数值。你应该以列的形式列出以避免混淆。

2—计算所有 X 值的总和，并计算所有 Y 值的总和。

3—计算每个 X 值的平方，并计算每个 Y 值的平方。

4—计算 XY 的总和。

将这些值代入公式 5.2 的等式中：

$$r_{XY}=\frac{(10\times247)-(54\times43)}{\sqrt{[(10\times320)-54^2][(10\times201)-43^2]}} \tag{5.2}$$

好啦，你会在公式 5.3 中看到答案：

$$r_{XY}=\frac{148}{213.83}=0.692 \tag{5.3}$$

现在，使用 Excel 的 CORREL 函数



使用 Excel 计算两个变量的相关系数。

1.在工作表中录入具体数值，如图 5.1 所示。

2.选择你想输入函数的单元格。在这个例子中我们将计算相关的函数放入单元格 B12。

3.单击单元格 B12，按照如下的形式输入 CORREL 函数。

=CORREL(A2:A11,B2:B11)

接着按回车键。

或者使用公式→其他函数→统计→CORREL 以及我们在 1A 中学习的技术，在单元格 B12 插入 CORREL 函数。

4.如图 5.2 所示，计算相关系数所得数值回到了单元格 B12。注意，在图 5.2 的公

式栏中可以看到 CORREL 的完整表达，在单元格 B12 中，你可以看到计算所得数值 0.692。

	A	B
1	X	Y
2	2	3
3	4	2
4	5	6
5	6	5
6	4	3
7	7	6
8	8	5
9	5	4
10	6	4
11	7	5

图 5.1 计算相关系数的数据

B12 =CORREL(A2:A11,B2:B11)

	A	B	C	D
1	X	Y		
2	2	3		
3	4	2		
4	5	6		
5	6	5		
6	4	3		
7	7	6		
8	8	5		
9	5	4		
10	6	4		
11	7	5		
12		0.69213		

图 5.2 使用 CORREL 函数计算相关系数

更多 Excel

PEARSON 函数也是应该了解的非常好用的函数，它计算的是皮尔逊积距相关系数。提示一下，PEARSON 函数和 CORREL 函数是同义的，计算的结果相同。

定义一个数值范围真的比输入一系列特定的单元格引用集更便利。因此，你可以不用输入 A2:A11，而是定义范围 Correct *X* 和 Correct *Y*，然后使用定义的范围。你可以复习 1A 中关于范围定义的内容。

相关的图示表达：散点图

一种非常简单、直观的表示相关的方式是建立**散点图**（scatterplot），或**散布图**（scattergram）。这是数据集在坐标轴上分布形成的简单图示。

我们依据下面的步骤建立如图 5.3 所示的散布图，所用数据是在上面已经计算了相关系数的 10 对数值。



1—画出 *x* 轴和 *y* 轴。一般来说，变量 *X* 在横轴，变量 *Y* 在纵轴。使用制图纸会让这个步骤更容易。

2—依据你知道的数据标出数轴的值域范围。例如，在我们的案例中变量 *X* 的值域范围是 2~8，因此我们在 *X* 轴标出的值域范围是 0~9。值域范围标注得高一些或低一些没有影响——只要你留出数值出现的空间。变量 *Y* 的值域范围是 2~6，因此我们在 *Y* 轴标出的值域范围是 0~9。这类带标签的数轴有时能使完成的散点图更容易理解。

3—最后，对应每对数值（例如图 5.3 中所示的 2 和 3）我们在图中标注一点，标注的地方是 2 对应 *x* 轴、3 对应 *y* 轴。就如在图5.3中所看到的，点代表**数据点**（data point），是两个值的交叉点。

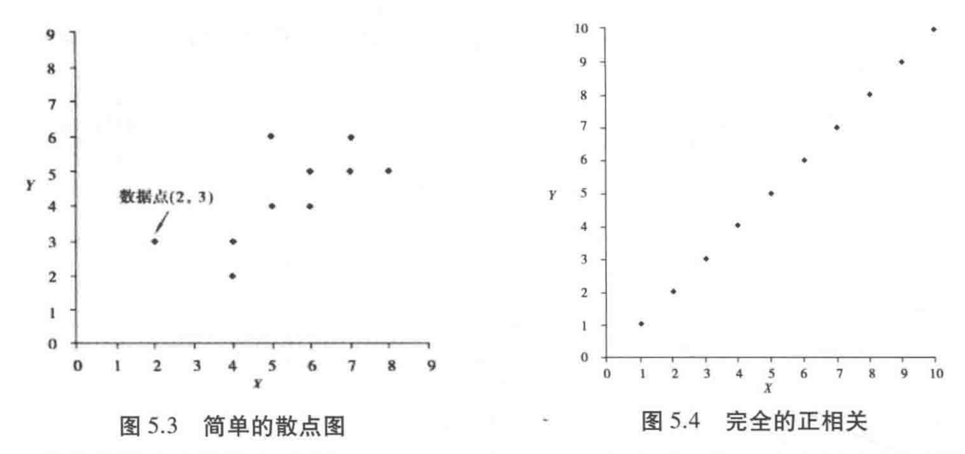


图 5.3 简单的散点图

图 5.4 完全的正相关

绘出了所有的数据点之后，我们所绘的这幅图就能告诉我们两个变量之间的关系。首先，数据点集合的形状表明了相关是正向的还是负向的。

如果一组数据点形成的点集是从 x 轴和 y 轴的左下角到右上角，就会出现正向的斜率。如果一组数据点形成的点集是从 x 轴和 y 轴的左上角到右下角，就会出现负向的斜率。

图 5.4—图 5.6 给出了各种不同相关的示例，从中你能看到一组数据点如何反映相关系数的强度和方向。

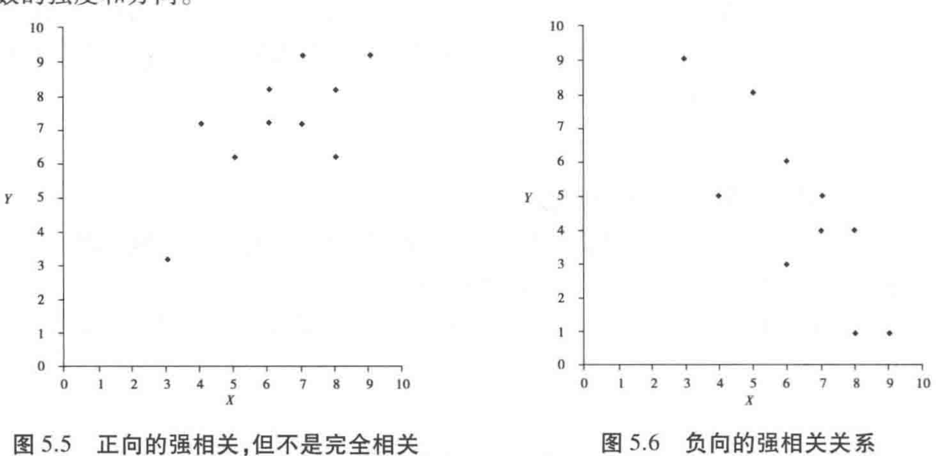


图 5.5 正向的强相关，但不是完全相关

图 5.6 负向的强相关关系

图 5.4 所示是完全正相关，$r_{XY}=1.00$，而且所有的数据点排成一条正向斜率的直线。

如果是完全负相关，相关系数值就是-1.00，而且数据点也会排成一条直线，但是直线是从图的左上角到右下角。换句话说，连接数据点的直线的斜率是负向的。

不要期望在行为科学或社会科学中能找到两个变量是完全相关的。也就是说，如果两个变量完全相关，它们就共享所有的特征。想想你的同学，你是否认为这些不同的人共享的一个特征和他们的另一个特征完全相关？很可能不是。实际上接近 0.70 和 0.80 的 r 值可能就是你看到的最大值了。

图 5.5 中你看到的是正向强相关的散点图，但不是完全相关，相关系数 $r_{XY}=0.70$。要注意的是数据点沿着正向的斜率排列，不是完全相关。

现在我们向你展示负向的强相关，如图 5.6 所示的相关关系，相关系数 $r_{XY}=-0.82$。要注意的是，数据点如何沿着负向的斜率从左上角向右下角排列。

这就是不同的相关类型看起来的样子，而且你可以依据数据点组合的方式大致判断相关的一般强度和方向。

使用 Excel 建立散点图

我们在第 4 章花费了很多时间和精力学习建立图表并调整其中的要素，但没学散点图（在本章学习）。我们已经知道散点图是什么了，现在来学习如何建立一个散点图。我们还是用图 5.1 所示的原始数据。只要单击几下就行了，非常神奇！

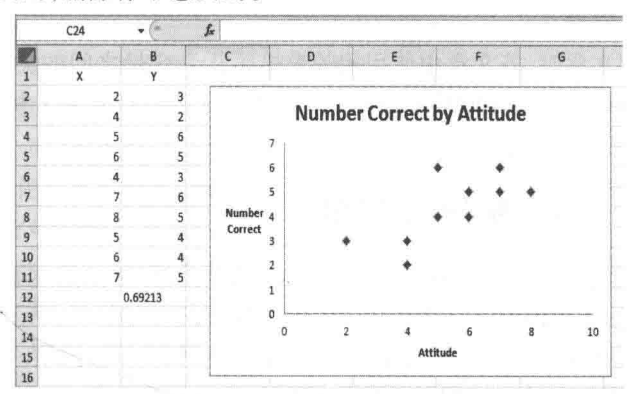


1—将单元格 A2 到 B11 的所有数据（不包括列标题）都选中。

2—单击［插入］按钮，在图表选项中单击［散点图］的下拉菜单，选择你想建立的散点图。我们单击第一个选项，图 5.7 中你可以看到已经完成的散点图（有一些小的调整）。

这看起来非常酷。

图 5.7 显示的是散点图的一种形式，我们调整了图的大小和背景，命名了两个数轴，赋予整个图表名称。这非常容易完成，只是需要多练习。记住操作过程中保存工作表，这样你就能保留你想要的，删除你不想要的。



	A	B
1	X	Y
2	2	3
3	4	2
4	5	6
5	6	5
6	4	3
7	7	6
8	8	5
9	5	4
10	6	4
11	7	5
12		0.69213

图 5.7 单击三次建立散点图



不是所有的相关都是用直线表明 X 值和 Y 值关系的**线性相关**（linear correlation）。相关关系可能不是线性的，而且也可能不是由直线反映的。就如年龄和记忆力之间的相关。在少年时期，相关可能是很强的正相关——儿童的年龄越大，他们的记忆力越好。到了青年和中年时期，没有太多的变化或相关程度变化不大，因为大多数青年人和中年人保持了良好的记忆力。但是在老年时期，记忆开始消退，而且在老年时期记忆力和年龄的增长是负相关关系。如果你将这些集中起来考虑，你会发现记忆力和年龄之间的相关看起来像曲线，也就是记忆力先增长，然后保持一定水平，接着就下降。这就是曲线相关，而且有时候曲线是对相关关系的最好描述。

相关系数集：相关矩阵

如果你有两个以上的变量怎么办？如何说明相关？如下所示的**相关矩阵**（correlation matrix）是非常简单和有效的解决方法。

	收入	教育	态度	选举
收入	1.00	0.35	−0.19	0.51
教育		1.00	−0.21	0.43
态度			1.00	0.55
选举				1.00

就如你看到的，矩阵中有四个变量：收入水平、教育水平、选举态度以及选举或投票的参与水平（取值范围是1~5）。

对应每一对变量都有一个相关系数。例如收入水平和教育之间的相关系数是0.35（也就是说，随着收入的增长，教育水平也在提高）。类似的，收入水平和选举态度的相关系数是-0.19，意思是说收入水平越高，这个人越可能不参加选举。

在这样的矩阵中，总是有4！/(4-2)！2！个相关系数，也就是说在4个变量中一次选择2个，总共会有6个相关系数。因为变量和它们自身完全相关（也就是沿着对角线的值都是1.00），而且因为收入和选举之间的相关系数与选举和收入之间的相关系数相同，矩阵建立了自身的镜像。

你在阅读期刊文章时，若文章中使用相关来描述几个变量之间的关系，你会看到很多这样的矩阵。你会在本章后面的部分看到，我们使用分析工具库来分析两个以上的变量之间的相关时，也会建立一个矩阵。

更多Excel——Excel中的相关矩阵

如果你有探究的欲望，并且想增强应用Excel的能力，你可以建立我们上面所示的矩阵，只要你在合适的单元格输入合适的CORREL函数形式，如图5.8所示。这里，我们同时插入公式中的定义名称（回顾1A中命名范围的内容）和CORREL函数。

但是（非常重要的但是），我们是在向你展示单元格中的公式而不仅仅是结果。我们如何完成这项任务？请记住，你可以看到公式和函数，而不仅仅是公式和函数呈现的结果，选中一个单元格或者整个工作表，接着使用Ctrl+`组合键，就会在公式和函数的结果与公式和函数之间实现转换。建立相关矩阵的数据会在四列中出现，每一列代表四个变量中的一个（收入、教育、态度、选举），相关的函数会在这些列的下端显示。

使用奇妙的分析工具库计算相关系数

如果你认为CORREL函数很奇妙，是因为你“还没见识过更厉害的”（谢谢Al Jolson）。虽然CORREL函数非常有用，但是分析工具库中的相关工具更好用。

我们还是使用图5.8所示的数据，按照下面的步骤来完成。

	A	B	C	D	E
1		Income	Education	Attitude	Vote
2		74190	13	1	1
3		80931	12	3	2
4		81314	11	4	2
5		73089	11	5	2
6		62023	11	3	2
7		61217	10	4	2
8		84526	11	5	1
9		87251	11	4	1
10		62659	12	5	2
11		76450	10	6	2
12		70512	12	7	2
13		78858	9	6	1
14		78628	13	7	1
15		86212	14	8	2
16		74962	9	8	2
17		58828	11	9	4
18		61471	10	8	5
19		78621	12	7	5
20		60071	9	8	4
21					
22		Income	Education	Attitude	Vote
23	Income	=CORREL(Income,Income)	=CORREL(Income,Education)	=CORREL(Income,Attitude)	=CORREL(Income,Vote)
24	Education		=CORREL(Education,Education)	=CORREL(Attitude,Education)	=CORREL(Vote,Education)
25	Attitude			=CORREL(Attitude,Attitude)	=CORREL(Attitude,Vote)
26	Vote				=CORREL(Vote,Vote)

图 5.8　使用 Ctrl+` 组合键显示各个单元格中的 CORREL 函数以创建相关矩阵



1.单击单元格 B22 作为结果返回的地址。

2.单击[数据]→[分析],你会看到数据分析对话框。是否还需要快速回顾如何使用? 见 1B 小节。

3.单击[相关系数](Correlation),接着单击[确定],你会看到如图 5.9 所示的相关系数对话框。

4.在输入区域框中,输入你想用 Excel 计算相关系数的数据范围。一定要将数据列的标签(将在分析结果中显示)包括进来。如图 5.10 所示,我们想用于分析的数据是单元格 B1 到 E20 的数据。

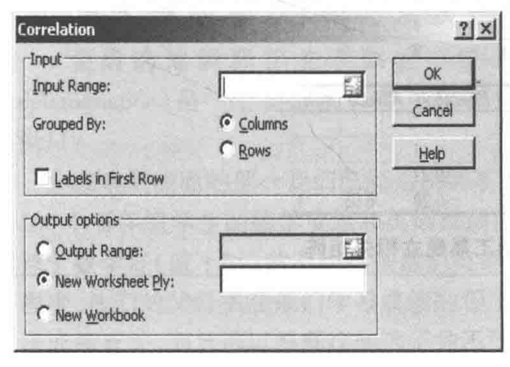


图 5.9　相关系数对话框

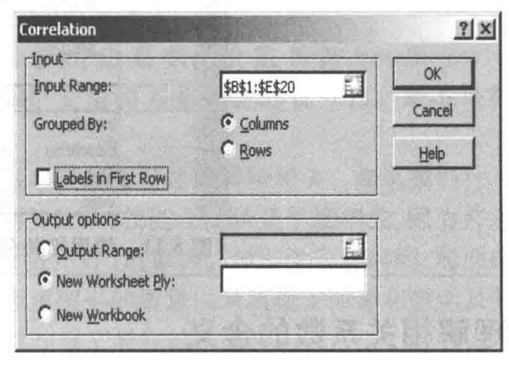


图 5.10　在相关对话框的输入区域中输入信息

5.选中[标志位于第一行](Labels in First Row)复选框。

6.在输出选项中单击输出区域选项。输入你想让 Excel 呈现分析结果的单元格位置。在这个案例中,我们选择 $B: $22。

7.单击[确定],结果就出来啦！图 5.11 所示结果很不错,你稍做一些改动就可以得到这样的效果。真的只需要很短的时间。

更多 Excel

要记住是 Excel 以默认的形式建立了相关矩阵。输入的数据与其他数据类似,你可以按照你的需要来处理,就如你在图 5.11 所见,我们已经对数据格式进行了微小的调整使得数据更简单明了,小数位进行了四舍五入,改变了总的呈现形式。

	Income	Education	Attitude	Age
Income	1.00			
Education	0.35	1.00		
Attitude	−0.19	−0.21	1.00	
Age	−0.51	−0.20	0.55	1.00

当使用分析工具库的工具时,它呈现的是结果而不是公式,因此在单元格中的是数字,而不是公式或者函数。你可以按照你的想法来进行调整。

	A	B	C	D	E	F
1		Income	Educatio	Attitude	Vote	
2		$74,190	13	1	1	
3		$80,931	12	3	2	
4		$81,314	11	4	2	
5		$73,089	11	5	2	
6		$62,023	11	3	2	
7		$61,217	10	4	2	
8		$84,526	11	5	1	
9		$87,251	11	4	1	
10		$62,659	12	5	2	
11		$76,450	10	6	2	
12		$70,512	12	7	2	
13		$78,858	9	6	1	
14		$78,628	13	7	1	
15		$86,212	14	8	2	
16		$74,962	9	8	2	
17		$58,828	11	9	4	
18		$61,471	10	8	5	
19		$78,621	12	7	5	
20		$60,071	9	8	4	
21						
22			Income	Education	Attitude	Vote
23		Income	1			
24		Education	0.35	1		
25		Attitude	-0.19	-0.21	1	
26		Vote	-0.51	-0.20	0.55	1

图 5.11 使用相关分析工具建立相关矩阵

理解相关系数的含义

现在我们已经有了变量间关系的量化指标,而且我们知道相关系数的值越大(不论正负号),相关关系就越强。但是相关的程度并不直接等于结果值,那么我们如何解释相关系数值,并使之成为更有意义的关系指标?

有几种方式可以来理解简单相关系数 r_{XY}。



你们中的一些人可能已经知道相关系数要检验其显著性，或者是其可能发生的概率水平，我们会在第 8 章学习。那是解释相关系数的另一种有用的方式，但是就现在而言，我们接下来学习的是简单的更具有描述性的方式。一旦了解了这些，你就为更复杂的学习内容做好了准备。

使用经验规则

也许最容易的（但不是信息量最高的）解释相关系数值的方式就是看相关系数的大小，并使用表 5.2 中的信息。

表 5.2　解释相关系数

相关系数的大小	一般解释
0.8~1.0	非常强的相关
0.6~0.8	强相关
0.4~0.6	中度相关
0.2~0.4	弱相关
0.0~0.2	弱相关或无关

因此，如果两个变量之间的相关系数是 0.5，你可以肯定地得出结论，变量间的关系是中度相关——不是强相关，也肯定不是没有任何共同特征的弱相关。

这种浏览式方法非常适合快速评价两个变量之间关系的强度，就如研究报告中的描述部分。但是，因为经验规则确实依赖主观判断（什么是“强相关”或“弱相关”），我们应该选择更精确的方法。这就是我们将要看到的内容。

而且，请注意，表 5.2 中有重叠。例如相关系数.4 既可以是弱相关，也可以是中等相关。数字不能说明所有问题，而是该由你（研究者）来做到。

决定性的努力：相关系数平方

更精确地解释相关系数的方法是计算决定系数。**决定系数**（coefficient of determination）是一个变量的方差可以被另一个变量的方差解释的百分比。是不是很拗口？

在本章前面的部分我们已经指出共享某些特征的变量之间如何相关。如果我们计算 100 个五年级学生的数学成绩和英语成绩的相关关系，我们会发现是中度相关，因为许多孩子数学好（或不好）的原因也是他们英语好（或不好）的原因。他们学习的时间、聪明的程度、他们的父母关注他们学习成绩的程度、家里图书的数量以及其他方面都和数学及英语成绩有关，而且可以解释这些孩子的不同（这也是变异性的来源）。

这两个变量共享的特征越多，它们就越相关。这两个变量也共享变异性，即孩子们之间存在差别的原因。总之，学习越努力、越聪明的孩子成绩越好。计算决定系数是为了确定一个变量的方差可以被另一个变量的方差解释的程度，决定系数就是相关系数的平方。

例如，如果平均成绩和学习时间之间的相关系数是 0.70（或者 $r_{GPA \cdot time}=0.70$），那么决定

系数——由 r^2 表示，是 0.7^2 或 0.49。这意味着平均成绩方差的 49%可以被学习时间的方差解释。相关越强，则越多的方差可以被解释（这很有道理）。这两个变量共享的特征越多（例如好的学习习惯、课堂能够学到的知识、没有压力），一个变量可以更多地解释另一个变量表示出的信息。

但是，如果 49%的方差可以被解释，就意味着 51%不能被解释——因此即使对相关系数是 0.70 的强相关来说，也存在很多无法解释的原因导致变量间的变化差异。不能解释的方差的数值就是**不相关系数**（coefficient of alienation）（也叫作**非决定系数**，coefficient of nondetermination）。这仅仅是不能被 X 解释的 Y 的方差的数量。

如何形象地表示共享方差的想法？你会在图 5.12 中看到一个相关系数、对应的决定系数，以及表示两个变量共享多少方差的图。每个图中灰色区域越大（两个变量共享的方差越大），这两个变量就越相关。

相关系数	决定系数	变量 X		变量 Y
$r_{XY}=0$	$r_{XY}^2=0$		共享 0%	
$r_{XY}=0.5$	$r_{XY}^2=0.25$ 或 25%			共享 25%
$r_{XY}=0.9$	$r_{XY}^2=0.81$ 或 81%			共享 81%

图 5.12 变量如何共享方差以及相应的相关

- 第一个图中两个圆没有接触。它们没有接触是因为它们没有任何共享的部分。相关系数为 0。
- 第二个图中两个圆有重叠。相关系数是 0.5（而且 $r_{XY}^2=0.25$），两个变量共享 25%的方差。
- 最后，第三个图中的两个圆几乎是一个覆盖于另一个之上。几乎是完全相关，相关系数 $r_{XY}=0.90$（$r_{XY}^2=0.81$），两个变量共享 81%的方差。

冰激凌吃得越多……犯罪率就越高（关联与因果关系）

这是计算、理解和解释相关系数时需要注意的真正重要的事。

想象一下，在美国中西部的一个小镇发现了一个不符合逻辑的现象。地方警察局局长发现冰激凌消费量越多，犯罪率就越高。这很简单，如果你测量这两个变量，你会发现这两个变量的相关关系是正向的，也就是人们吃的冰激凌越多，犯罪率就越高。就像你可能预期的一样，他们吃的冰激凌越少，犯罪率就越低。这个警察局长很是困扰，直到他回想起他在大学选修的统计学课程。

他的这个困惑最后变成了一个"哈哈"！他想"非常容易"，这两个变量一定是共享什么或彼此之间有共同之处。要记住，一定存在什么变量同时和冰激凌消费水平、犯罪率水平相关。你能猜到是什么吗？

室外温度是他们共有的特征。当室外气温变暖，如在夏天，就会有更多犯罪（白天更长，人们多开窗户，等等）。而因为天气变暖，人们更享受吃冰激凌的乐趣。相对地，在又长又黑暗的寒冬岁月，冰激凌的消费就减少，同时犯罪也越少。

乔·鲍勃(Joe Bob)近来被选举为市议员,他知道了这个发现并且有了一个很好的想法,或者至少他认为他的选民会喜欢这个想法(记住,他忽略了大学开设的统计学课程)。为什么不在夏天这几个月限制冰激凌的消费量,以使犯罪率下降?听起来很合理,对吧?得啦,不用进一步检验就能知道,这个结论根本没有意义。

这里有一个简单的原则,那就是相关表示两个或更多变量之间存在关联;相关并不是因果关系。换句话说,仅仅因为冰激凌消费水平和犯罪率一起增长(或一起下降)并不意味着一个变量的变化会导致另一个变量的变化。

例如,如果我们将镇子里所有商店的冰激凌拿走而且不再贩售,你认为犯罪率会下降吗?当然不会,连这样想都是十分荒谬的。但是奇怪的是相关经常被解释成这样(就好像真的存在因果关系),而社会科学和行为科学的复杂问题就会因为误解沦落为微不足道的琐事。长头发和嬉皮士与越南战争有关吗?当然不是。犯罪数量的上升和更有效、更安全的轿车有关吗?当然不是。但是它们都是同时发生,而且建立了关联的假象。

其他重要的相关系数

评价变量的方式很多。例如定类变量的属性是类别的,如种族(白人或黑人)或政党背景(无党派或共和党)。或者你测量收入和年龄,两者都是定距水平的变量,因为作为它们建立基础的连续统具有相等的间距。当你继续研究,你可能遇到发生在不同测量水平之间的数据的相关。而要计算这样的相关你需要一些特殊的技术。表 5.3 汇总了这些不同的技术及其区别。

表 5.3 相关系数展示,选哪一个?

测量水平和例子		相关类型	要计算的相关
变量 *X*	变量 *Y*		
定类的(选举偏好,如共和党或民主党)	定类的(性别,如男性或女性)	卡方系数	选举偏好和性别之间的相关
定类的(社会阶层,如上层、中层或下层)	定序的(高中毕业班的排序)	等级二列相关系数	社会阶级和在高中学校排序之间的相关
定类的(家庭结构,如双亲家庭或单亲家庭)	定距的(平均成绩)	点二列相关系数	家庭结构和平均成绩之间的相关
定序的(转换为排序的身高)	定序的(转化为排序的体重)	斯皮尔曼等级相关系数	身高和体重之间的相关
定距的(解决的问题数量)	定距的(以年计的年龄)	皮尔逊相关系数	解决的问题数量和以年计的年龄之间的相关

小　结

表明一个变量和另一个变量如何相关以及它们共有什么特征的想法非常有影响力，也是非常有用的描述统计（也在推论统计中使用）。要记住的是，相关表明的关系仅仅是关联的，但不是因果的，你能够理解这项统计如何就变量间的关系提供有意义的信息，以及一个变量在另一个变量变化时如何变化或保持不变。

练　习

1.使用这些数据回答问题 1a 和 1b。这些数据的数据集名称为第 5 章数据集 2。

（20 个问题之中）回答正确的总数	对考试的态度（总分是 100）
17	94
13	73
12	59
15	80
16	93
14	85
16	66
16	79
18	77
19	91

a.手动计算皮尔逊积距相关系数，并记下所有的过程。

b.手动建立这 10 对数据的散点图。依据散点图是否可以预测相关是正向的还是负向的？为什么？

2.使用下面的数据回答问题 2a 和 2b。

（游完 50 米）速度	力量（推举的重量级）
21.6	135
23.4	213
26.5	243
25.5	167
20.8	120
19.5	134
20.9	209
18.7	176
29.8	156
28.7	177

a.使用计算器或 Excel 软件计算皮尔逊相关系数。

b.使用相关系数从最弱到最强的值域范围来解释数据，并计算决定系数。主观分析和 r^2 值比较结果如何？

3.计算下面的数据的相关系数并进行解释。数据集是第5章数据集2。我们研究医生培训的年限和他们的手术成功度之间的关系。我们使用特定量表来评价医生成功与否,分数是1-10,1表示成功度高,10表示成功度最低。

4.使用第5章数据集3和分析工具库计算洗手效果(Wash)、每天洗手次数(Number)和医院感染数量(Infect)之间的相关系数。洗手效果使用1-5测量水平,1表示最好。对于结果你会做出怎样的解释?

5.两个变量之间的决定系数是0.64,回答下面的问题:

a.皮尔逊相关系数是多少?

b.是否强相关?

c.这两个变量的关系中不可解释的方差是多少?

6.看表5.3。你选择哪一种相关系数检验种族(以类别来定义)和政治立场(民主党、共和党和无党派)之间的关系?俱乐部成员身份与高中平均成绩之间的关系呢?解释你为什么做出这样的选择。

7.如果变量之间相关,为什么会相关?如果两个变量之间强相关,我们是否可以得出这样的结论,即一个变量的变化会引起另一个变量的变化?为什么可以或为什么不可以?

8.假定你是地方诊所的负责人,你正在地方学区调查,来了解辍学率和青少年怀孕之间的关系。一个地方议员拿到了研究报告,向公众做了如下的声明:"显然孩子们在学校待的时间越长,他们怀孕的可能性就会越低。"这项声明有什么问题?

9.下面是一组数据,计算皮尔逊相关系数,并解释计算的结果。

预算增加	12个月内课堂效果提高
7%	11%
3%	14%
5%	13%
7%	26%
2%	8%
1%	3%
5%	6%
4%	12%
4%	11%

10.依据相关关系强度对下面的相关系数进行排序(先列出相关强度最弱的)。

+0.71　　+0.36　　−0.45　　+0.47　　−0.62

6 这就是真相——理解信度和效度

本章你会学到什么 ☺☺☺

- 什么是信度和效度以及它们为什么重要
- 这是统计学课程中学习测量必不可少的内容
- 基本的测量尺度
- 如何计算和解释不同类型的信度系数
- 如何计算和解释不同类型的效度系数

信度和效度介绍

社会福利以及其他领域的专家认为美国50多万寄养儿童的存在是个严重的问题。主要问题之一是:在出生家庭仍然在他们的生活中扮演重要角色的情况下,寄养儿童如何适应暂时的寄养家庭。

索尼亚·J.莱瑟斯(Sonya J.Leathers)在研究亲生父母的经常性拜访与寄养儿童对寄养家庭和养父母的忠诚度之间的关系时考察了这个问题。在199个青少年样本中她发现出生家庭的经常性拜访的确带来了冲突,她建议采取一些有助于减少这些冲突的干预措施。

为了完成研究,她使用了许多不同的依赖变量,她通过儿童症状量表和访谈收集数据。在其他方面,她做得相当正确的就是选择了已经建好且具有可接受信度和效度水平的测量工具。不是每一个研究者都进行的一步,却是本章我们关注的重点。

如果想了解更多,请查阅莱塞斯(Leathers,S.)2003年发表在《家庭关系》(*Family Relations*)第52期53-63页的文章"寄养儿童的父母拜访、冲突的忠诚以及情绪和行为问题(Parental visiting, conflicting allegiances, and emotional and behavioral problems among foster children.)"。

为什么学习测量

这是非常好的问题。毕竟你选修的是统计学课程,而且到现在为止,学习的内容都是统计学的内容。现在你所面对的问题似乎属于检验和测量课程。那么,这部分内容在统计学书中有什么作用?

这是你应该提出的一个很好的问题。为什么?《爱上统计学:Excel》到目前所讲的内

容都与收集、分析和解释数据有关。我们即将开始学习如何分析和解释数据。在开始学习这些技能之前,我们要确保数据正是你想要的——数据显示的就是你想知道的内容。换句话说,如果你在研究贫困问题,你要确保用于评估贫困的测量工具都能发挥作用。或者你研究中年男性的侵略性,你要保证评估侵略性的任何工具能发挥作用。

还有更多的好消息:如果你想继续深造,并且想选修检验和测量课程,这一章的介绍会极大地促进你理解这个领域和将要学习的那些课程。

为了保证整个数据收集过程以及那些代表一定意义的数据都是有用的,你首先要保证用于收集数据的工具能发挥作用。本章将要回答的基本问题是"我怎么知道我每一次使用的检验、量表和工具等都能发挥作用?"(这就是信度)以及"我怎么知道我每一次使用的检验、量表和工具等能够测量我想测量的内容"(这就是效度)。

更多 Excel

在第 5 章我们着力介绍了相关系数及其应用,我们还会在本章继续讨论(在第 16 章会有更多讨论)。你之前学的使用 Excel 的内容,都可以在这里应用。Excel 没有专门的叫作"信度计数器"的工具,但是可以使用 CORREL 函数和分析工具库中的相关(Correlation)工具。

任何研究人员都会告诉你建立检验工具的信度和效度的重要性,不论是对消费行为的简单观察工具还是测量复杂的心理状态(如依恋)的工具。此外,还有一个很好的原因。如果你用来收集数据的工具是不可信或无效的,那么任何假设检验的结果都不会有结论。如果你不能确定检验能够完成应该完成的工作并且保持一致性,那么你怎么知道你得到的显著性结果是由于有问题的检验工具,而不是在零假设为真的情况下实际上拒绝了零假设(第一类错误)? 你是否想要一个"清白的"零假设检验? 如果是,那么现在就开始关注信度和效度。

关于测量尺度

在我们对信度和效度进行更多的讨论之前,我们首先要讨论不同类型的**测量尺度**(scales of measurement)。什么是测量? 就是依据一定的规则给观察结果分配一定的数值——很简单。最后得到的就是我们待会儿要定义的不同的尺度,而且一个观察结果可以是我们有兴趣测量的任何事物,如头发的颜色、性别、考试成绩或身高。

测量的尺度或规则是测量观察结果的特定水平。每一个水平都有特定的属性特征集。测量尺度以四种形式(或类型)出现:定类、定序、定距和定比。现在简单地介绍并举例说明四种不同类型的测量尺度。

玫瑰的种称:定类测量水平

定类测量水平(nominal level of measurement)是以观察结果的属性特征定义,也就是观察结果只适合一个而且唯一的一个分类或层级。例如,性别是定类变量(男性或女性);种族(高加索人或非裔美国人)和政党背景(共和党、民主党或无党派人士)也是定类变量。

定类水平的变量是“名称(name)”(拉丁语中是 nominal)，而且是准确水平最低的测量。定类测量水平的各个类别相互排斥，例如，政党背景不能同时是民主党和共和党。而一株玫瑰也不能既属于波旁玫瑰种(不开玩笑，就是波旁王朝那个波旁)，又属于多花蔷薇种。

我喜欢任何次序：定序测量水平

定序测量水平(ordinal level of measurement)的“序”表示次序，而且被测量的事物按照它们的属性特征排序。例如一份工作的应征者的次序，拉斯的次序是 1，谢尔顿的次序是 2，汉娜的次序是 3，那么这就是定序安排。我们不知道就这个尺度拉斯相对于谢尔顿比谢尔顿相对于汉娜高多少。我们只是知道次序 1 比次序 2 和次序 3 要靠前，但是不知道靠前多少。

1+1=2：定距测量水平

现在我们取得了一些进展。当我们谈到**定距测量水平**(interval level of measurement)时，是指检验或评估工具是基于某种连续统，这样我们就可以讨论一个较高的成绩比较低的成绩高多少。例如，你的词汇测试成绩是 10 个单词正确，就是 5 个单词正确的两倍。定距尺度的一个显著特征是尺度上的每个间距都相等。10 个单词正确比 8 个单词正确多 2 个，而 8 个又比 5 个单词正确多 3 个。

一个人可能一无所有吗？定比测量水平

你可能有些迷惑。**定比测量水平**(ratio level of measurement)的评估工具的特征是测量尺度中有绝对零值。这意味着被测量的特质完全不存在。那么让人迷惑的是什么？我们测量的观察结果是否可能是没有一丁点儿被测量的特质存在的？在一些学科中可能存在这种情况。例如，在物理和生物学中可以有不存在属性特征的情况，如绝对零值(没有分子运动)或零光程。在社会和行为科学中，这就比较让人迷惑了。即使你的拼写成绩是 0 或者答错了 IQ 测试的每一个题目，也并不意味着你的拼写能力为 0 或者智力为 0 吧？

总　之……

这些测量尺度或规则表示观察结果在特定水平被测量。而且，我们可以这样说：

- 任何测量结果都能够归属到四个测量尺度中的某一个。
- 测量尺度有一定的层次，从最不准确的定类尺度到最准确的定比尺度。
- 测量尺度“越高”，收集的数据越准确，并且数据包含的细节和信息越多。例如，了解到一些人富裕一些人贫困可能已经足够(这是定类或分类的区别)，但是准确地知道每一个人收入的多少(定距或定比)会更好。一旦我们知道每一个人收入的所有信息，就很容易对其进行简单的贫富划分。
- 最后，比较高的测量尺度包含所有在其之下的测量尺度的特性，如定距尺度包含定类尺度和定序尺度的特性。例如你知道熊的攻击平均值是 350，也知道这比老虎(老虎的攻击平均值是 250)好 100，同样你也就知道熊比老虎好(但不知道好多少)，而且熊与老虎之间存在不同(但是不知道到底怎样不同)。

总而言之，请看下表。表中列出了你选择正确的测量水平需要的内容。

	特征			
尺度	绝对零值	等距点	排序数据	类别数据
定比	✓	✓	✓	✓
定距		✓	✓	✓
定序			✓	✓
定类				✓

信度——再做一次直到得到正确的值

信度(reliability)很好解释,就是一个测试或者其他任何测量工具对事物的测量可以保持一致性。如果你在特定的处理方案之前进行人格测试,那么四个月之后再进行的相同测试是否可信?这是关于信度的问题之一,有不同的信度类型。定义信度之后我们会对每一类型的信度进行更多的介绍。

检验值——真本事还是运气

你参加这个课程的考试,得到一个成绩,可能是 89 分(这很好)或 65 分(回去好好看书!)。这样的考试包含几个不同要素,包括观察值(observed score,你实际得到的考试成绩,如 89 分或 65 分)和真实值(true score,真实的,100%准确反映你对学习内容的真实掌握)。我们不能直接测量真实的成绩,因为这个值是个体所拥有的实际特征或特性在理论上的数据反映。

测验和测量之间存在许多的争议,而真实值肯定具备其存在的理由,下面做出解释。我们刚刚定义的真实值是与观察对象的某种性质或者特点关联的非常真实的真实值。这样定义到现在还都没问题。但是还存在另外的观点。一些心理学测量专家(以做测验和测量为生的人)认为真实值与调查的内容是否真的被反映之间没有关系。而且,真实值是测验者测验无数次之后可以得到的平均值,表示一个给定测验的理论上的典型的测验能力水平。一般可以预期典型的能力水平反映要调查的内容,但是这是另外一个问题(也就是效度)。这里的区别是如果一项测验可以使得一个人在平均意义上持续得到相同的分数,那么这项测验就是可信的,不论测验的是什么。事实上,一个完全可信的测验也不能得出一个与要测量的内容完全关联的数值,就如"你真的知道什么"。

真实值和观察值为什么不相同?如果考试(以及相应的观察值)是对测量内容的完美反映(我们是指绝对的完美),它们就会是相同的。

但是人们不能保持长胜,错误总是在不经意间。墨菲定律(Murphy's law)告诉我们世界并不完美。因此,你看到的是观察值可能非常接近真实值,但是极少完全相同。而差异——就如你在这里看到的——包含在将要介绍的误差中。

观察值=真实值+误差值

误差?的确是误差。例如,我们假定某个学生的统计学成绩是 80 分,但是他的真实成绩

（我们永远不会真的知道，只是理论上可以假定）是89分。这意味着9分的差异（就是**误差值**）来自误差，或者另一个导致个人考试成绩与100%真实成绩不同的原因。

这样的误差的来源是什么？可能是考试的房间太热而让你犯困，这肯定对你的考试成绩有影响。或者是由于你没有按计划为考试复习。这两个例子反映考试的环境或状况而不是被测量的特征，对吧？

我们的工作是尽可能减少误差，例如改善考试环境并且保证你能得到足够的睡眠。减少误差、增加信度，这样观察值才能和真实值更匹配。

误差越小就越可信——就是这么简单。

信度的不同类型

信度有多种不同的类型，现在介绍最重要也是最常用的四种，总结在表6.1中。

表6.1　信度的不同类型，何时使用，如何计算以及它们的意义

信度的类型	何时使用	如何计算	举例说明所得结果的含义
再测信度	你想知道一个测试在不同时间是否可信	计算时期1和时期2相同测试的两次值之间的相关系数	关于青少年认同形成的邦佐（Bonzo）测试是可信的
复本信度	你想知道一个测试的几个复本是否可信或者是否是等价的	计算一个复本测试的值与相同内容的另一个复本的测试（不是完全相同的测试）的测试值之间的相关系数	人格测试的两个复本是等价的，而且表现出复本信度
内在一致性信度	你想知道一个测试的项目是否评价一个而且只评价一个维度	每一个项目的得分与总得分之间的相关系数	SMART创造性测试的所有项目评价相同结构
评分者信度	你想知道对一个观察结果的评价是否具有一致性	检验不同评分者一致结论的百分比	足球运动员最佳着装评价的不同评分者信度是0.91，表示不同裁判的一致程度很高

再测信度（前测-后测信度）

再测信度（test-retest reliability）用于检验一个测试在不同时期是否可信。例如，你想建立检验不同类型职业项目的选择偏好测试。

你在六月份进行了测试，接着在九月份进行了相同的测试（保持相同很重要）。那么你计算两次测试的得分集（记住是相同的人进行了两次测试）是否相关，也就是进行了信度的测试。再测信度是检验不同时间的变化或差异所必需的信度。

你必须确保你测量的内容是以可信的方式测量的，这样你得到的测试结果才可能与每个个体每个时点的测量值更接近。

计算再测信度。下面给出MVE（管理职业教育测试）在时期1和时期2的部分测试值。我们的目标在于计算皮尔逊相关系数作为这一测量工具的再测信度。

测试者编号	时期 1 得分	时期 2 得分
1	54	56
2	67	77
3	67	87
4	83	89
5	87	89
6	89	90
7	84	87
8	90	92
9	98	99
10	65	76

这个过程的第一步和最后一步是计算皮尔逊积矩相关系数(复习第 5 章相关内容),这个值等于 0.90。

$$r_{Time1 \cdot Time2} = 0.90 \tag{6.1}$$

我们会简单地解释这个值。

复本信度

复本信度(parallel forms reliability)用于检验相同测试的不同复本的等价性和相似性。

例如,你在研究记忆,有的研究内容是看一眼给定的 10 个单词并尽可能记住,然后在 20 秒钟记忆、10 秒钟休息之后背诵这些单词。因为这些研究的进行需要两天的时间,也涉及一些记忆技能的培训,你需要符合研究任务要求的另一组单词,而且很明显不能是相同的那些词。因此,你建立另一个词汇表,而且希望这个词汇表和上一个类似。在这个案例中,你希望不同复本的一致性很高——测试内容相同,只是形式不同。

计算复本信度。下面给出来自 IRMT(自主记忆测试)的复本 A 和复本 B 的部分数据。我们的目标在于计算皮尔逊相关系数作为测量工具的复本信度。

测试者编号	复本 A 得分	复本 B 得分
1	4	5
2	5	6
3	3	5
4	6	6
5	7	7
6	5	6
7	6	7
8	4	8
9	3	7
10	3	7

这个过程的第一步和最后一步是计算皮尔逊积矩相关系数(复习第 5 章相关内容),这个值等于

$$r_{FormA \cdot FormB} = 0.13 \tag{6.2}$$

我们会简单地解释这个值。

内在一致性信度

内在一致性信度(internal consistency reliability)与之前介绍的两种类型非常不同。内在一致性信度用于确定测试中的项目是否彼此一致,都只表示一个维度、一个结构或一个关注的领域。

例如,你在进行对不同类型的医疗保健的态度测试,而且你想确保这一组5个项目都是针对态度而不是其他的测试。你会看到(一个测试群体)每一个项目的得分,然后确定个体得分是否和总体得分相关。你预期在某个项目(例如,我喜欢我的HMO——美国健康保护组织)上得分高的测试者会在其他项目上得分低(例如,我不想在医疗保健上花钱),而且这对所有的测试者都是一样的。

计算 α 系数或克隆巴赫系数(Cronbach's)。下面给出10个测试者在5个态度项目测试(我爱HMO测试)上的部分样本数据,其中每个项目的得分在1(非常不同意)到5(非常同意)之间。

测试者编号	项目1	项目2	项目3	项目4	项目5
1	3	5	1	4	1
2	4	4	3	5	3
3	3	4	4	4	4
4	3	3	5	2	1
5	3	4	5	4	3
6	4	5	5	3	2
7	2	5	5	3	4
8	3	4	4	2	4
9	3	5	4	4	3
10	3	3	2	3	2



当你计算克隆巴赫系数(以Lee Cronbach命名)时,实际上就是计算每个测试者在每个项目上的得分和总得分之间的相关系数,并与所有单个项目得分的变异性比较。计算的逻辑是每个总分很高的测试者在每一个项目上的得分应该也很高(若总分40的测试的每一个项目的得分为5,5,3,5,3,4,4,2,4,5),而每个总分很低的测试者在每一个项目上的得分应该也很低(若总分40的测试的每一个项目的得分为5,1,5,1,5,5,1,5,5,1,5,1,就是缺少一致性或者不是一维的)。

下面是计算克隆巴赫系数的公式:

$$\alpha = \left(\frac{k}{k-1}\right)\left(\frac{s_y^2 - \sum s_i^2}{s_y^2}\right) \tag{6.3}$$

其中:k 表示项目的个数,

s_y^2 表示观察值的方差,

$\sum s_i^2$ 表示每一个项目的方差的总和。

下面的数据与前面给出的数据是相同的,还加上了完成上面的等式需要计算的值(观察值的方差,或 s_y^2,以及每个项目的方差的总和,或 $\sum s_i^2$)。

测试者编号	项目1	项目2	项目3	项目4	项目5	总分
1	3	5	1	4	1	14
2	4	4	3	5	3	19
3	3	4	4	4	4	19
4	3	3	5	2	1	14
5	3	4	5	4	3	19
6	4	5	5	3	2	19
7	2	5	5	3	4	19
8	3	4	4	2	4	17
9	3	5	4	4	3	19
10	3	3	2	3	2	13
						$s_y^2 = 6.4$
观察值方差	0.32	0.62	1.96	0.93	1.34	$\sum s_i^2 = 5.17$

将具体的数值代入这个公式，你就会得到下面这个等式：

$$\alpha = \left(\frac{5}{5-1}\right)\left(\frac{6.40 - 5.17}{6.40}\right) = 0.24 \tag{6.4}$$

你会发现 α 系数为.24，现在你已完成了内在一致性信度计算（之后对这个值进行解释）。不好意思——这里没有使用酷酷的 Excel 工具。



如果我们告诉你还有许多其他类型的内在一致性信度，你不会感到惊讶吧？这对内在一致性的测量来说是事实。不仅有 α 系数，也有折半信度，斯皮尔曼-布朗（Spearman-Brown），库德-理查逊（Kuder-Richardson）20 和 21（KR_{20} 与 KR_{21}）以及其他进行这类测试的系数，它们只是以不同的方式检验测试工具的一维性。

评分者信度

评分者信度（interrater reliability）是两个评分者对观察结果判断的一致程度的测量。

例如，你的研究兴趣是银行工作人员和潜在顾客在交易过程中相互作用的类型，你在现场观察（在单面镜后观察）银行工作人员在接受了新的或高级的顾客关系课程之后是否会对潜在顾客发生微笑等友善的行为。你的工作是记录每 10 秒内银行工作人员是否展示课程所教的三种不同的行为类型——微笑、坐着时身体前倾或者用手指出某个要点。每次看到三种行为中的任何一种出现，就在你的积分表上标注"/"。如果没有观察到任何一种，就标注为"—"。

作为这个过程的一部分，而且为了保证你的记录是可信的，你需要知道不同的观察者记录这些行为发生的一致水平。记录的形式越类似，不同的评分者的一致水平和信度就越高。

计算评分者信度。在这个案例中，真正重要的问题是在两分钟内每 10 秒钟（或者是 12 个 10 秒钟内）对顾客的友好行为是否发生。因此，我们要看的是在分成 12 个 10 秒钟的 2 分钟的时间框架内记录的一致性。记录表中一个斜线（/）代表着行为发生，一个横线（—）表示行为没有发生。

	时期⟶	1	2	3	4	5	6	7	8	9	10	11	12
记录者 1	戴夫	/	—	/	/	/	—	/	/	—	—	/	/
记录者 2	莫瑞恩	/	—	/	/	/	—	/	/	—	/	—	/

在 12 个时期内(存在 12 个可能的一致),戴夫与莫瑞恩一致认为有 10 个时期行为发生或没有发生(时期 1、2、3、4、5、6、7、8、9 和 12),而对 2 个时期行为发生情况有争议(时期 10 和 11),也就是说有 10 个记录一致,而剩下 2 个记录不一致。

评分者信度可以使用下面的简单公式计算：

$$评分者信度=\frac{一致的数量}{可能一致的数量} \tag{6.5}$$

代入具体的数值就会得到下面的等式：

$$评分者信度=\frac{10}{12}=0.833 \tag{6.6}$$

最后得到的评分者信度系数是 0.833。

多大才是大——解释信度系数

现在我们开始讨论信度系数。你是否还记得第 5 章所学的对相关系数的解释？这与解释信度系数几乎是相同的,只有很小的差异。

我们只需要确定两个方面,也就是：

- 信度系数是正向的,而不是负向的
- 信度系数一般都很大(在 0.00 和+1.00 之间)

如果你不能建立信度,怎么办

建立某个测试的信度并不容易,而且出错的原因可能不是你没有好好努力工作。如果测试是不可信的该怎么办？

下面的几个方面需要谨记在心。记住信度是表示观察值产生的误差是多少的一个函数。误差越小,信度就越高。

- 在进行测试时,确保对所有题项的设置都是标准化的。
- 增加项目或者观察的数量,因为从你观察的总体行为中得到的样本越大,样本的代表性和信度越高。特别是对成绩的测试来说是这样的。
- 去掉有歧义的项目,因为一些人会对它做出一种反应,另外一些人会做出不同的反应,无论他们的知识、能力水平或个性特征如何。
- 特别是对成绩测试(如拼写考试、历史考试)来说,应当注意调整测试的难度,因为任何测试太难或太容易都不能准确地反映一个人的水平。
- 减少外部事件的影响,并且将答题指导标准化,因此如果特定事件在临近测试时发生,如狂欢节或毕业,就要推迟测试进行的时间。

还有一点

建立具备合理心理学支持的测验工具的第一步是建立信度（我们只需多花一些时间就行）。为什么，如果测试或测量工具不可信，或者没有一致性，而且不能在不同的时间进行相同的测试，那么测量到什么还重要吗（这是效度问题）？

如果拼写入门的 KACAS（Kids Are Cool at Spelling，孩子对拼写很在行）测试的前三个项目是：

16+12=？

21+13=？

41+33=？

这肯定是信度很高的测试，但是肯定不是有效的测试。现在我们对信度已经有了很好的了解，接下来开始介绍效度。

效度——什么是真实的？

效度（validity），最简单地说，就是表示工具能够测量想测内容的程度。有效的测试是测量到了理应测量的内容。如果成绩测试的是历史知识，那么就测量历史知识。如果智商测试是测量建立者定义的智商内容，那么就是测量这个。

效度的不同类型

就如信度有不同的类型，效度也有不同的类型，我们介绍最重要也是最常用的三种类型。这几种类型都总结在表 6.2 中。

表 6.2 效度的不同类型、何时使用、如何计算以及它们的意义

效度的类型	何时使用	如何计算	举例说明所得结果的含义
内容效度	你想知道一个特定主题的项目样本是否能够反映项目总体。	请教专家，让专家判断测试的项目是否反映将要测量的主题的项目总体。	我的统计学课程的每周测试可以评价每章的内容。
准则效度	你想知道测试成绩是否和其他标准系统相关，这个标准表明被测试者具备某个领域的能力。	计算测试成绩和其他有效的测量之间的相关系数，并评价相同的一组能力。	研究显示烹调技能的 EAT 测试成绩与结束烹饪学习后两年内成为主厨是相关的（预测效度的一个案例）。
建构效度	你想知道测试是否测量一些基本的心理结构。	计算测试成绩和反映测试设计的结构的理论结果之间的相关系数。	这是真实的——参与身体接触和危险性运动的男性的侵略性 TEST 测试的得分较高。

内容效度

内容效度(content validity)就是测试项目能代表计划测试要测量的总体项目的性质。内容效度常用于成绩测试(例如从一年级的拼写测试到学术能力测试)。

建立内容效度。建立内容效度实际上非常容易。你需要做的就是确定具体的合作专家。例如,我要设计物理入门测试,我会寻找合适的物理专家(也许是地方高中的教师或者大学教授物理学的教授),而且我会说"你好,艾伯特,你看这 100 个多项选择题能不能准确地反映我期望的入门课的学生掌握的所有可能的主题和概念?"

我可能告诉艾伯特或艾伯塔具体的主题是什么,然后他或她看了具体题目之后就会判断出这些题目是否符合我建立的标准(代表入门课程中的全部内容)。如果答案是肯定的,我就完成了设计(至少就现在而言);如果答案是否定的,就要重新开始建立新的题目或者修正现存的题目。

准则效度

准则效度(criterion validity)是评价测试能否在现在和将来的情境中反映一组能力。如果准则是发生在现在,我们就讨论**同步效度**(concurrent validity)。如果准则是发生在未来,我们就讨论**预测效度**(predictive validity)。对于准则效度的应用,不需要同时建立同步效度和预测效度,按照测试目的的需要选择适用的那个就行了。

建立同步效度。例如你受雇于世界烹饪学院设计测量烹饪技能的工具。烹饪培训的部分内容与直接的知识有关(例如什么是奶油面粉糊? 这属于测试的内容)。

于是,你建立了你认为能够很好测量烹饪技能的测试表,现在你想建立同步效度水平。为完成这项工作你设计了烹饪量表(COOK scale),将每个裁判都使用的判断标准(如品相、清洁等)以 5 分制的项目形式集中在一起。作为标准(这是关键),你还有另一个裁判组,他们将学生的整体技能进行从 1 到 10 的等级排序。接着只需要计算烹饪量表得分和裁判的排序之间的相关系数。如果效度系数(简单相关系数)很高,你的设计就很好;否则就需要重新开始。

建立预测效度。例如,我们知道烹饪学校 10 年来发展得很好,而你不仅仅对厨师厨艺的好坏感兴趣(这是你刚才建立的同步效度的部分),而且对预测效度也感兴趣。现在准则已经从当下的得分(裁判的给分)转变为未来的得分。

现在我们的目的是建立一个测试,可预测一个厨师 10 年后的成功。要建立烹饪测试的预测效度,你要回顾并找到结束烹饪课程后 10 年仍然从事烹饪工作的毕业生,对他们进行测试。这里使用的标准是他们成功的水平,而且你使用的测量指标是:(a)他们是否有自己的餐馆;(b)餐馆经营是否超过 1 年(因为新餐馆经营的失败率在第 1 年就超过 80%)。基本原理是如果餐馆经营超过一年,那么主厨的厨艺一定很好。

要完成这项练习,就要计算值为 1(如果餐馆经营超过 1 年并且为毕业生所有)的成功得分与之前的(10 年前)烹饪得分之间的相关系数。较高的相关系数表明具有预测效度,较低的相关系数表明缺乏预测效度。

建构效度

建构效度(construct validity)是最有趣也是最难建立的效度,因为建构效度是基于测试或测量工具背后的基本结构或概念的。

你可能从初级心理学教材中了解到,结构指一群相关的变量。例如,侵略性是一个结构(包括不恰当的碰触、暴力、缺乏成功的社会交往等),智力、母婴依恋和希望等同样都是结构。而且要记住的是这些结构来自某种研究者假定的理论视角。例如,他或她认为有侵略倾向的男性比没有侵略倾向的男性更容易陷入与权威的矛盾中。

建立建构效度。好,现在你拿着好斗性测试(FIGHT test)(基于侵略性的)量表,它是由一系列依据你对侵略性结构构成的理论概括编制的项目构成的观察工具。你从犯罪学的文献中了解到具有侵略性的男性会比其他人更容易做出某种类型的行为,例如,他们更多地与人争论,他们更易发生肢体冲突(如推挤),犯下更多的暴力罪行,而且在人际关系方面更少能取得成功。好斗性量表包括的项目描述了不同的行为,其中一些行为在理论上和侵略性行为有关,另一些无关。一旦好斗性量表完成,你就应该检验结果来确定好斗性量表的得分与你预测的行为类型(如卷入犯罪的水平或人际关系的性质等)之间是否相关,以及是否与理论上应当无关的行为类型(如没有遭遇家庭暴力或完成高中、大学学业等)不相关。如果你预期相关的项目的相关系数很高,而且预期无关的项目的相关系数很低,那你所应用的好斗性量表是有效的(不过也有可能是你设计的那些不是评价侵略性要素的项目起了作用)。恭喜你!

如果不能建立效度怎么办

这个问题很难回答,特别是因为效度类型是如此之多。

总之,如果你没有得到你想要的效度证明,这是由于你的测试没有测到预期的内容。如果是成绩测试,而且你寻求满意的内容效度水平,那么你可能要重新设计测试的问题以确保这些问题与专家认为应该测试的问题保持一致。

如果你关注准则效度,你可能需要重新检验测试的项目的性质,并且设置你期望人们的回答与你选定的准则之间有多大的相关程度的问题。

最后,如果你寻求建构效度但不能找到——最好重新认真思考支持你建立测试的理论基础。那么要么是我们关于侵略性的定义和模型是错的,要么是现有知识还需要一些批判性的思考。

最后的建议

测量是一件相当酷的事——能激起人们的兴趣,而且在注重责任的时代,每个人都想知道学生、股票经纪人、社会福利机构项目等的进展。

由于这类强烈的、激增的兴趣,撰写学年论文或毕业论文的大学生和研究生都十分倾向于给自己的论文方案设计工具。

但要注意的是,听起来很好的想法可能导致的是一场灾难。建立任何工具的信度和效度都需要几年的紧张工作。而轻信的或无怀疑精神的人想建立新的工具检验新的假设可能使这个过程更糟糕。这意味着在检验新假设的同时,还得确保新工具发挥作用。

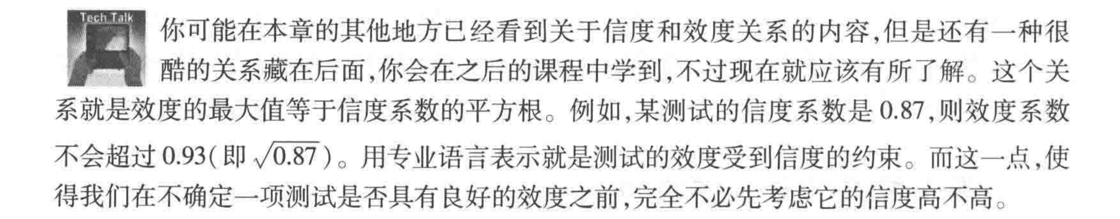
如果你在做自己的原创性研究，如毕业论文或学位论文，要确保找到的测量已经建立了信度和效度。这样的话你才可以将精力主要用在检验假设上，而不是将精力浪费在发展工具这项工作上——这是一项职业，本身就需要专门的工具测评。想要有一个好的开始，试试查阅比勒什心理测量研究所(Buros Institute of Mental Measurements)的网站，可以从那里获得在线资料。

信度和效度：亲密的表亲关系

现在我们暂时止步并回顾学习这一章的原因之一。

这章很重要，因为你需要了解用于测量的工具的信度和效度。为什么？如果工具既不可信又无效，你的实验结果也就让人怀疑。

就如本章之前提到的，你的测试可能是可信却无效的。但是，你的有效的测试却不可能没有信度。为什么？不论一个测试测试的是什么都可以重复进行(这就是信度)，但是可能仍然没有测量到应该测量的内容(这就是效度)。但是，如果一个测试确实测量了要测量的内容，那么这个测试重复操作的结果必然是一致的。



你可能在本章的其他地方已经看到关于信度和效度关系的内容，但是还有一种很酷的关系藏在后面，你会在之后的课程中学到，不过现在就应该有所了解。这个关系就是效度的最大值等于信度系数的平方根。例如，某测试的信度系数是 0.87，则效度系数不会超过 0.93(即 $\sqrt{0.87}$)。用专业语言表示就是测试的效度受到信度的约束。而这一点，使得我们在不确定一项测试是否具有良好的效度之前，完全不必先考虑它的信度高不高。

小　结

这就是统计学课程里的测量内容。再提一次，任何统计学的应用都是围绕一些测量的结果进行的。就如你需要基本的统计学使得数据具有意义，你也需要基本的测量信息使得对行为的评分、考试成绩、等级排序及其他的评价有意义。

练　习

1. 去图书馆查找 5 篇你的专业领域的给出信度和效度数据报告的文章，并且讨论所用测量工具的结果。确定要建立的信度类型和效度类型，并且考虑你是否认为这个水平可接受。如果不能接受，如何改进？
2. 提供建立再测信度和复本信度的案例。
3. 下面是 15 个成年人在 6 月和 8 月进行的相同的长时记忆测试的得分。你使用什么信度类型建立信度，这个测试是否可信？

6月测试得分	8月测试得分
8	9
4	6
5	5
6	6
1	3
7	7
6	8
8	9
7	6
6	7
7	8
8	6
9	9
10	9

4.假设你现在建立了一个测量职业偏好(你想要成为什么人)的工具,你需要在学生参加职业项目的这一年进行多次测试。你需要从两次测量的数据(从第6章数据集1获得数据)——一次测量在秋季,一次测量在春季,评估这项测量的再测信度。这是否是可信的测量?为什么?

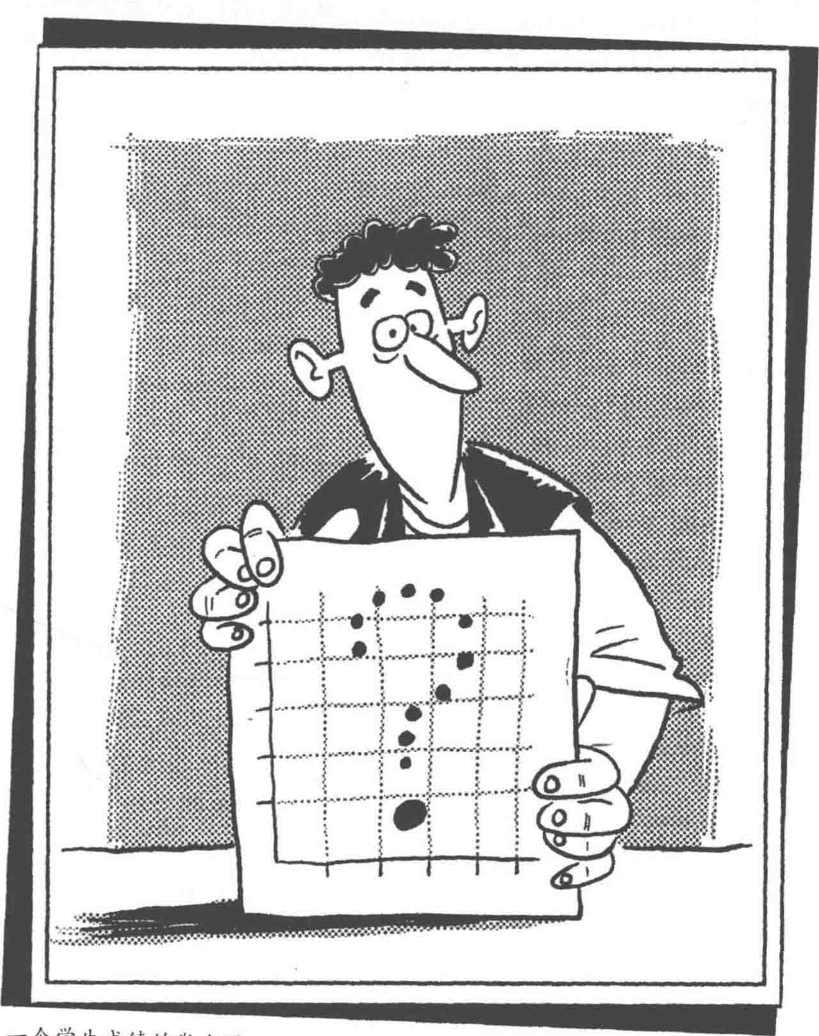
5.为什么一项测试可能是可信的但却是无效的?为什么一项测试只有是可信的才可能是有效的?

6.在检验任何实验假设时,为什么用于测量结果的测试工具的信度和效度很重要?

7.描述两种准则效度类型的差异,并分别给出一个案例进行说明。

第Ⅲ部分

抓住那些有趣又有利的机会



一个学生成绩的散点图。

到目前为止你已了解了很多,接下来又会学什么呢？首先,你已经具备了如何描述一组数据的特征以及数据分布如何区别的坚实基础。这是你在《爱上统计学:Excel》第2章、第3章和第4章学到的内容。在第5章你学习了如何使用相关工具描述变量间的关系。在第6章你学习了如何计算信度和效度,判断测评工具的有效性。

现在是加大赌注开始玩真儿的时候了。在《爱上统计学:Excel》第Ⅲ部分,第7章向你介绍假设检验的重要性和性质,包括假设检验是什么、有几种不同的类型、假设的功能以及假设为什么和如何被检验的深度讨论。

接着,在第8章我们通过讨论正态曲线说明概率的所有重要方面,以及作为概率基础的基本原则。作为统计学的一部分,概率能帮助我们定义某类事件(如一次考试中的特定值)发生的可能性。我们将使用正态曲线作为这些讨论的基础,你会看到一个数据分布中任何数值或事件的发生都具有一定的概率水平。

带着乐趣学习了概率和正态曲线之后,我们准备在第Ⅳ部分开始更进一步的讨论,主要是依据假设检验和概率论的应用来检验有关变量间关系的特定问题。从现在开始只会越来越好。

你和假设:检验你的问题 7

本章你会学到什么 ☺☺☺☺（今天就别计划晚上出去玩了！）

- 样本和总体之间的差异
- 零假设和研究假设的重要性
- 判断一个好假设的标准

也许你想成为一个科学家

你可能已经在其他的课程中听到了假设这个词,甚至不得不为了在其他课上所做的研究建立一个假设,或者你在期刊文章中已经看到过一两次。如果是这样,你可能对假设是什么已经有了很好的认识。对于你们中不熟悉这个词的人来说,**假设**(hypothesis)基本上是“学术猜测”。假设最重要的角色是表示一般问题的陈述,或在最初的时候促使人们去探索研究的问题。

> **更多 Excel**
>
> 第7章是本书中不需要使用 Excel 来学习基本统计的一章。这一章主要是一些想法。理解这些想法很重要,我们会在随后的章节中使用 Excel(从下一章开始)来说明其中的某些想法。

这就是为什么花费精力和时间去建立一个简要而清晰的研究问题是如此重要的原因。研究问题是建立假设的指导,相应的假设决定你用于检验假设以及回答问题的技术。

因此,一个好的假设将问题陈述或研究问题转换为更适合于检验的形式。这种形式就叫作假设。我们将在本章后面的部分讨论如何才能建立一个好的假设。在此之前,我们的注意力会转向样本和总体的差异。这是重要的区别,因为假设检验是用于样本,然后才将结论一般化到更大的总体。接着将注意力转向假设的两个主要类型(零假设和研究假设)。现在首先让我们给在前面章节用到的一些简单的术语作正式的定义。

样本和总体

作为一名好的科学家,你可能会说方法 A 比方法 B 好,这是永远、一直而且对宇宙所有人都正确的吗?实际上如果你依据方法 A 和方法 B 相应的优点进行足够的研究,并检验了足够的人数,总有一天你能这样说。但是不要太激动,因为你不可能如此确信地说。

这会花费太多的钱($$$)和太多的时间(用所有的人)来做那些研究,另外这是不必要的。可行的是你从总体中选择一个代表性样本,并且检验有关方法 A 和方法 B 的假设。

几乎所有的科学家都受到没有足够的时间和研究经费的限制,所以最好的策略是从一个较大总体中选取一部分,并在这个较小的群体中进行研究。在这种情况下,较大的群体作为**总体**(population),从这个总体中选择的较小群体就是**样本**(sample)。

Tech Talk 测量样本和总体特征近似程度的量数叫作**抽样误差**(sampling error)。抽样误差基本上就是样本统计值和总体参数之间的差异。抽样误差越大,抽样过程中的精确性越低,且应用样本中的发现真实反映总体中的预期发现就更加困难。就如数据分布有不同的变异性量数,样本量数和总体量数之间的差异也有变异性量数。

样本应该以这样的方式从总体中选取,就是样本要尽可能和总体的特征匹配。目的就是使得样本尽可能地类似总体。保证这两个群体的类似的最重要目的是基于样本的研究结论可以一般化到总体。如果样本确实代表了总体,就可以说研究结论具有很高的一般化水平。

一般化水平高是好研究的重要特点,因为这意味着进行研究所花费的时间和精力以及金钱不仅仅是对研究的群体有意义,而且对整个群体有意义。

将"代表性"与"大量的"联系起来是很容易的。要记住的是,一个代表性样本要比一个大的样本更重要(人们经常认为越大越好——可能只适用于感恩节)。例如,在一个样本中有很多参与者令人印象深刻,但是如果这些参与者不能代表总体,研究的价值就很小。

零假设

好吧,我们已经从总体中选择了一个样本来检验研究假设,我们接下来要建立**零假设**(null hypothesis)。

零假设非常有趣。如果它能讲话,一定会说类似这样的话:"我表示你们正在研究的两个变量无关。"换句话说,零假设可以由下面这些取自主流社会和行为科学期刊的真实的(简短的)案例说明。为了保护个人隐私我们做了相应更改。

- 9 年级学生的 ABC 记忆考试的平均成绩和 12 年级学生的平均成绩没有差异。
- 社区长期照料老人的效果和家庭长期照料老人的效果没有差异。
- 反应时间和问题解决能力无关。
- 白人家庭和黑人家庭提供给孩子的与学校活动有关的支持在数量上没有差异。

这四个假设的共同之处在于他们都宣称两个或多个事物之间是等同的或没有关系的。

零假设的目的

零假设的基本目的是什么？零假设既是研究起点也是测量实际的研究结果的基准。现在让我们更详细地讨论每一个目的。

首先,零假设是研究的起点,因为在没有其他信息的情况下零假设就被看作可接受的

真实状态。例如,我们上面给出的第一个零假设:

9 年级学生的 ABC 记忆考试的平均成绩和 12 年级学生的平均成绩没有差异。

如果对 9 年级和 12 年级学生的记忆技能没有更多了解,就没有理由相信这两个群体之间存在差异,对吧? 如果对变量间的关系没有任何了解,你能做的最好的就是去猜测。这就要承担一定的风险。你可能对一个群体为什么比另一个群体做得好做出猜测,但是你没有先验的(在事实之前)证明,那么除了假定他们是相同的之外还有什么选择?

"不存在某种关系"作为研究的起点是整个假设的重点。换句话说,除非你能证明存在差异,否则你只能假定没有差异。而无差异或无关的陈述正是零假设的所有内容。

进一步讲,如果这两个群体之间存在任何差异,你必须假定这些差异是出于偶然,这是对任何两个群体之间在任何变量上的差异的最有吸引力的解释——存在偶然性! 这是对的,如果没有其他信息,偶然性总是对观察到的群体之间的差异或变量之间的关系的最可能和最有吸引力的解释。偶然性解释我们不能解释的差异。你可能已经把偶然性看作是在游戏机上赢得 5 000 美元大奖的几率,但是我们讨论的偶然性完全是另外的"事物",它掩盖事实,甚至使得理解变量间关系的"真实"性质变得更困难。

例如,你选取了一组橄榄球队员和一组足球队员来比较他们的跑步速度。但是要考虑所有我们不知道的可能导致速度差异的因素。谁知道是不是一些橄榄球队员进行了更多的练习,或者一些足球队员更强壮? 或者两组队员都接受了额外的训练? 更重要的是,也许测定他们速度的方式本身就有很大的偶然性;出问题的秒表或大风天可能导致了与真实速度无关的差异。作为好的研究者,我们的工作是解释观察到的差异时消除偶然性因素,并分析其他可能导致群体差异的因素,例如有目的的训练或营养计划,并分析这些因素如何影响速度。重点是如果我们发现群体间有差异且差异不是由于训练引起的,我们就很难将差异归结为偶然性之外的其他因素。

零假设的第二个目的是提供与观察到的结果进行比较的基准,进而分析是否是由于其他因素引起这些差异。零假设有助于定义观察到的群体间的差异范围是由偶然性引起(这是零假设的论点)还是由偶然性之外的因素(这可能是其他变量影响的结果,例如上个例子中的训练)引起。

许多研究暗含着零假设,也许你不能在报告和期刊文章中清楚地发现零假设的表述。但是你会发现明确表述的研究假设,这是我们接下来要关注的内容。

研究假设

零假设是变量间无关的陈述,而**研究假设**(research hypothesis)是变量间有关系的明确陈述。例如对于之前陈述的每一个零假设,都有对应的研究假设。注意我们说到相应的研究假设时用的表达方式,因为对任何一个零假设来说肯定存在不止一个研究假设。

- 9 年级学生的 ABC 记忆考试的平均成绩不同于 12 年级学生的平均成绩。
- 依据 Margolis 社会活动量表的测量,社区长期照料老人的效果不同于家庭长期照顾老人的效果。
- 反应时间越慢则问题解决能力越高,两者正相关。
- 白人家庭和黑人家庭提供给孩子教育活动方面的支持在数量上有差异。

这四个研究假设有一个共同的方面。它们都是不等价的陈述。它们假定变量间存在

一定的关系，而不是零假设假定的等价的关系。

不等价关系可以采取两种形式——有方向研究假设和无方向研究假设。如果研究假设假定不等价关系没有方向（例如“不同于”），假设就是无方向研究假设。如果研究假设假定不等价关系有方向（例如“多于”或“少于”），这个研究假设就是有方向研究假设。

无方向研究假设

无方向研究假设（nondirectional research hypothesis）反映群体间的差异，但是差异的方向是不确定的。

例如研究假设

9 年级学生的 ABC 记忆考试的平均成绩不同于 12 年级学生的平均成绩

是无向的，也就是两个群体间差异的方向不确定。这是一个研究假设，因为它陈述存在差异，但没有陈述差异的方向。

这里描述的假设是无方向研究假设，可以用下面的式子表示：

$$H_1: \overline{X}_9 \neq \overline{X}_{12} \tag{7.1}$$

其中：H_1 是表示第一个（可能有几个）研究假设的符号，

$\overline{X}_9$ 表示 9 年级学生样本的平均记忆成绩，

$\overline{X}_{12}$表示 12 年级学生样本的平均记忆成绩，

$\neq$表示“不等于”。

有方向研究假设

有方向研究假设（directional research hypothesis）反映群体间的差异，而且差异的方向是确定的。

例如研究假设

12 年级学生的 ABC 记忆考试的平均成绩比 9 年级学生的平均成绩高

是有向的，因为两个群体间差异的方向是确定的。一个被假设大于（不仅仅是不同）另一个。

其他两个有方向假设的案例是：

A 大于 B（或 A>B），或 B 大于 A（或 B>A）

这两个假设都表示不相等，且不相等是特定性质的（大于或小于）。上面描述的 12 年级学生的成绩比 9 年级学生的成绩好的假设是有方向研究假设，可以用下面的式子表示：

$$H_1: \overline{X}_{12} > \overline{X}_9 \tag{7.2}$$

其中：H_1 是表示第一个（可能有几个）研究假设的符号，

$\overline{X}_9$ 表示 9 年级学生样本的平均记忆成绩，

$\overline{X}_{12}$表示 12 年级学生样本的平均记忆成绩，

>表示“大于”。

研究假设的目的是什么?直接检验的研究假设是研究过程中的重要一步。通过比较检验的结果与随机预期的结果(也就是零假设),来确定这两个中哪一个是你观察到的群体间差异的更好解释。

表 7.1 是 4 个零假设和相应的无方向研究假设和有方向研究假设。

表 7.1 零假设和相应的研究假设

零假设	无方向研究假设	有方向研究假设
9 年级学生的 ABC 记忆考试的平均成绩和 12 年级学生的平均成绩没有差异。	12 年级学生的 ABC 记忆考试的成绩不同于 9 年级学生的成绩。	12 年级学生的 ABC 记忆考试的平均成绩高于 9 年级学生的平均成绩。
依据 Margolis 社会活动量表的测量,社区长期照料老人的效果和家庭长期照顾老人的效果没有差异。	依据 Margolis 社会活动量表的测量,社区长期照料老人的效果不同于家庭长期照顾老人的效果。	依据 Margolis 社会活动量表的测量,社区长期照料老人的效果高于家庭长期照顾老人的效果。
反应时间和问题解决能力无关。	反应时间和问题解决能力有关系。	反应时间和问题解决能力之间正相关。
白人家庭和黑人家庭提供给孩子的支持数量没有差异。	白人家庭提供给孩子的支持数量不同于黑人家庭提供的支持数量。	白人家庭提供给孩子的支持数量高于黑人家庭提供的支持数量。



讨论有方向和无方向假设的另一种方式就是讨论单尾和双尾检验。**单尾检验**(one-tailed test,反映有方向假设)假定了特定方向的差异,如我们假设群体 1 的得分比群体 2 高。**双尾检验**(two-tailed test,反映无方向假设)假定差异没有特定的方向。如果你想检验不同类型的假设(单尾和双尾的检验),建立拒绝或接受零假设的概率水平,这个区别就十分重要。在第 10 章会有更多的讨论,我保证。

研究假设和零假设的一些区别

除了零假设表示等价关系而研究假设表示不等价关系之外,这两类假设还有几个重要的不同之处。

第一,简短地回顾一下,两类假设的区别在于一个(零假设)表示两个变量之间没有关系(相等),而另一个(研究假设)表示两个变量之间有关系(不相等)。这是基本的区别。

第二,零假设总是对应总体,而研究假设总是对应样本。我们从一个较大的总体中选择一个样本。接着我们试图将样本的结论一般化到总体中。如果你还记得基本的哲学和逻辑原理(你已经上过了这些课,对吧?),就会知道从一个小群体(如样本)到一个大群体(如总体)的过程就是推论。

第三,因为总体不能直接进行检验(再说一次,这是不现实、不经济的,而且通常是不可能的),你不能百分之百肯定地说样本之间在某些变量上真的不存在区别。另外,你只能依据样本的研究假设的检验结论来做出间接推论。因此零假设只能被间接检验,而研究假设能够被直接检验。

第四，零假设常用希腊字母表达，而研究假设常用罗马字母表达。例如9年级学生的平均成绩等于12年级学生的平均成绩的零假设可以如下表示：

$$H_0:\overline{X}_9 = X_{12} \quad (7.3)$$

其中：H_0 表示零假设，

X_9 表示9年级学生总体的理论平均值，

X_{12} 表示12年级学生总体的理论平均值。

研究假设是12年级的样本的平均值高于9年级的样本的平均值，如公式7.2所示。

最后，因为你不能直接检验零假设，所以零假设是暗含的假设。而研究假设是明确的，且便于表达。这就是你在研究报告中很少看到零假设而几乎总是看到研究假设陈述的另一个原因。

好假设的标准是什么

现在你知道假设就是学术猜测——是进一步研究的起点。对所有的猜测来说，一开始其中的一些就比另一些好。我们得强调提出你想回答的问题多么重要，记住你提出的任何假设都是对你要问的原始问题的直接扩展。问题反映你个人的兴趣和动机，以及已完成的研究。了解这些之后，你可以使用下面的标准来确定在研究报告中看到的假设或者自己建立的假设是否为可接受的假设。

我们使用一个研究案例来说明这一点，这项研究是考察为工作到很晚的父母提供的放学后儿童照料对父母工作适应性的影响。下面是一个很好的假设：

> 让自己的孩子加入放学后儿童照料计划的父母一年之内耽误的工作日会减少，同时依据工作态度调查的测量，工作态度也会比没有让自己的孩子加入这个计划的父母更积极。

下面就是标准。

第一，一个好的假设一般是以陈述句的形式出现而不是以问题出现。在上面的案例中，没有提出“你是否认为父母和他们工作的公司会更好”这样的问题，这是因为如果陈述是明确、有力的，假设就更有效。

第二，一个好的假设提出变量间预期的关系。案例中使用的假设清楚地描述了放学后儿童照料、父母的态度和缺勤率之间的关系。这些变量会接受检验来考察一个变量（加入放学后儿童照料计划）对其他变量（缺勤率和态度）的影响。

注意到上面的标准中“预期的”一词吗？定义预期的关系是为了避免像钓鱼或乱放枪一样漫无目的地建立一些无用的关系。

Tech Talk

钓鱼方法就是你把线扔出去，然后抓住任何咬了线的东西。你搜集尽可能多的数据而不管研究兴趣是什么，或者你搜集的数据是否是科学调查的一部分。或者你把枪装满子弹，然后向任何移动的东西射击，你肯定能射中什么。问题是，你可能不想要你射中的，更糟糕的是你可能错过你想要的，最坏的是（如果可能）你可能不知道你射中的是什么。好的研究人员不是抓到或者射中什么就要什么，他们想要特定的结果。研究人员想得到这样的数据就需要开放性的问题和明确、有力且容易理解的假设。

第三，假设反映它们建立的理论和文献基础。就如在第1章看到的，科学家的成功很少是只归结于他们自己的努力工作。他们的成功通常是由于，或者说部分是由于走在他们前面并留下解释框架的科学家。一个好的假设要反映这一点，这样假设就和已存在的理论和文献连接在一起。在上面的案例中，我们假定有文献表明父母知道孩子在妥当的环境中得到照料就会感到更放心，然后就能在工作中表现得更好。了解这些就可以假设放学后照料孩子计划能提供父母所寻求的安心。相应地，父母就会集中精神工作而不是不断地打电话确定他们的孩子是否安全到家。

第四，假设应该简短并切中要点。你要使假设以陈述句的形式描述变量间的关系，并且尽可能直接和明确。越是切中要点，其他人（如你的论文答辩委员会成员）就越容易阅读你的研究、理解你的假设是什么以及重要的变量是什么。实际上，当人们阅读并评价研究（你会在后面的章节学到更多）时，他们中的大多数做的第一件事就是找到假设，从而对研究的一般目的和研究是如何进行的有了很好的了解。一个好的假设可以反映这两个方面。

第五，好的假设是可检验的假设。这意味着你可以实际地回答假设中包含的问题。你可以从上面的示例假设中看到让孩子加入儿童照料计划的父母和没有让孩子加入的父母之间的比较。态度或缺勤的天数是可测量的变量。两者都是合理的目标。态度可以通过工作态度调查表来测量，缺勤（的天数）很容易被精确地记录。如果建立的假设是"让孩子加入儿童照料计划的父母对工作的感觉更好"，想想还会有多少困难。即使你得到相同的信息，"感觉更好"这个模棱两可的词就会使得结果更难解释。

总之，假设应该：

- 以陈述句的形式表述，
- 假定变量间的关系，
- 反映假设建立的理论和文献基础，
- 简短并切中要点，
- 可检验。

如果假设满足这五个标准，你可以判断这个假设很好，足以继续进行研究来准确地检验作为假设来源的一般问题。

小　结

任何科学研究的中心要素是假设，而不同的假设类型（零假设和研究假设）有助于形成计划来回答我们的研究所提出的问题。零假设作为研究起点和比较基准的特点使得我们可以使用零假设来评价研究假设的可接受性。接下来我们继续学习如何实际检验零假设。

练　习

1. 去图书馆查找你感兴趣的领域的五篇实证研究论文（包含数据）。对于每一个研究，列出下面的内容：

   a. 零假设是什么（潜在的还是明确的陈述）？

   b. 研究假设是什么（潜在的还是明确的陈述）？

c.那些没有明确的或潜在的假设的文章又怎样呢？确认这些文章，并看看你是否可以建立一个研究假设。

2.你在图书馆的时候，在你感兴趣的领域选择两篇文章，简短地介绍样本以及如何从总体中选择样本。一定要介绍研究者是否做了大量的工作来选择样本，并且能够回答你的问题。

3.依据下面的研究问题，建立一个零假设，一个有方向研究假设，一个无方向研究假设。

a.注意力对教室里不专心听课行为的影响如何？

b.婚姻的质量和夫妇双方与他们兄弟姐妹间关系好坏之间的关系如何？

c.治疗厌食症的最好方式是什么？

d.儿童3岁前的早期干预项目对其6年级的阅读技能的影响如何？

4.回到问题1中找到的5个假设，并用本章最后部分提及的5个标准进行评价。

5.什么是零假设，零假设的一个重要目标是什么？零假设和研究假设的区别是什么？

6.为什么零假设假定变量之间没有关系？

7.现在做更刺激的练习。依据下面的内容提供一个零假设和一个研究假设，以及等式。

a.大学生运动员的食品支出和其他大学生的食品支出。

b.小白鼠和小棕鼠走出迷宫的平均时间。

c.治疗某种疾病的药品A和药品B的效果。

d.使用方法1和方法2完成任务的时间。

你的曲线是正态的吗——概率和概率的重要性 8

本章你会学到什么

- 为什么理解概率是理解统计学的基础
- 什么是正态曲线或钟形曲线，其特点是什么
- 如何计算和解释 z 值

为什么学习概率

现在你认为这是统计学课程了吧！好吧，就如你在这一章将学到的，学习概率是理解正态曲线（接下来会对这个有更多讨论）的基础，也是理解推论统计的基础。

为什么？首先，正态曲线给我们提供了理解任何可能结果（如一次考试中得到某个具体分数，或者投掷硬币时得到正面）的概率的基础。

其次，概率研究是决定我们在陈述特定的发现或结果是“真”时所具有的可信度的基础。或者更好的说法是，某个结果（如平均成绩）没有出现是由于偶然因素。例如，我们比较 A 群体（每星期参加 3 小时的额外游泳训练）和 B 群体（没有额外训练）。我们发现 A 群体的适应性测验和 B 群体不同。但是我们可以说这个差异是由于额外训练引起，或者是其他因素引起的吗？概率研究提供的工具允许我们确定差异是由于练习或其他因素引起的准确的可能性数值。

上一章我们花费在假设上的所有时间都是值得的。一旦我们把对零假设和研究假设的理解与概率基础的想法结合在一起，就可以讨论特定结果（由研究假设建立）出现的可能性。

正态曲线（或钟形曲线）

什么是正态曲线？**正态曲线**（normal curve）也叫作**钟型曲线**（bell shaped curve）或**钟形曲线**，就是具备三个特征的数据分布的形象表示。图 8.1 表明了这三个特征。

正态曲线表示均值、中位数和众数相等的数值分布。你可能还记得第 4 章的内容，也就是如果中位数和均值不同，那么数据分布就向某个方向倾斜。正态曲线没有偏度。正态曲线有一个很好的波峰（只有一个），而且波峰正好处于中间。

其次，正态曲线以均值为中心完全对称。如果沿着中心线将曲线对折，两边会完全重叠。两边是相等的。曲线的一半是另一半的镜像。

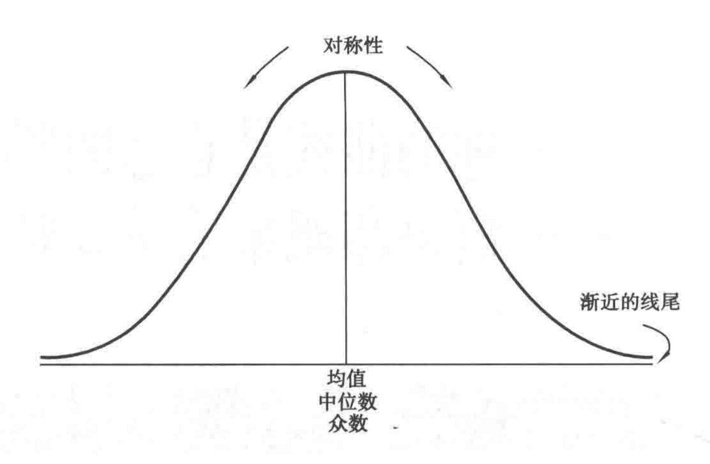


图 8.1 正态曲线或钟形曲线

最后（准备好学习新的内容），正态曲线的双尾是渐近的（asymptotic）——一个很陌生的词。这个词的含义是曲线的双尾越来越逼近横轴，但是永远不会与横轴相交。

正态曲线的钟形形状给了这个曲线另外一个名称，也就是钟形曲线。



作者在年轻的时候也想知道正态曲线的尾线如何逼近横轴或 x 轴却永不相交。试着做下面的实验。放置两支相距 1 厘米的铅笔，然后移近一半，他们就相距 1/2厘米，再移近 1/4 厘米，再移近 1/8 厘米。这两支铅笔会持续的接近，对吧？但是不会（永远不会）相交。正态曲线的尾线同样如此。尾线缓慢地接近曲线“栖居”其上的数轴，但是永远不会真正相交。

这一点为什么很重要？在本章后面的部分你将会了解，尾线从不相交的事实意味着可能得到的极值（在曲线的左侧或右侧）存在无限小的可能性。如果尾线和数轴相交，就得不到原本可能存在的极值了。

嘿，这不是正态曲线

我们希望你接下来的问题是，“但是有许多数据集的分布不是正态的或不是钟形的，对吧？”是的。但当我们处理大样本数据集（超过 30 个数据），并且重复地从总体中抽取样本时，曲线的数值就接近正态曲线的形状。这很重要，因为我们讨论的通过样本推论总体时所做的大多数工作都是基于这样的假设——总体中抽取的样本是正态分布的。

而且事实是大体上许多事物的分布特征就是我们所说的正态分布。也就是说大多事件或发生次数正好在数据分布的中间，而两端却较少，就如在图 8.2 中所看到，图 8.2 表示一般人群中 IQ 和身高的分布。

例如，一个群体中聪明人很少，处在群体底端的智力或认知能力很低的人也很少。大多数人正好处在曲线的中间，而我们移到曲线的尾线时人数减少。另一个案例是，一个群体中很高个的人相当少，很矮个的人也相当少，而大多数人处在身高分布的中间。在这两个案例中，智力和身高的分布接近正态分布。

相应地，正态曲线中极值范围内发生的每一个事件具有很小的发生概率。我们可以

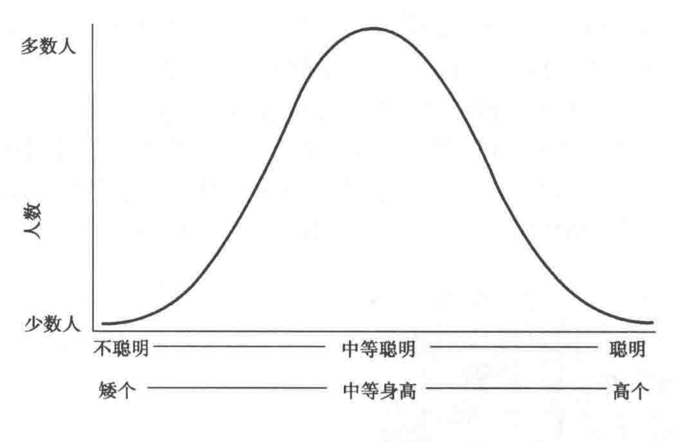


图 8.2 数值如何分布

很确信地说任何人(我们还不知道他们的身高)个子很高的几率不是很大。但是我们知道任何人处于平均身高的几率或者正好处于中间范围的几率很高。那些发生在正态曲线中间的事件的概率高于极值范围内发生的事件的概率。

中心极限定理

这里我们讨论曲线,曲线,以及更多的曲线,并且你现在可能已经有了这样的认识,即每个数值都有对应发生的概率。你也会认识到,这就是我们提出,什么是随机发生什么不是随机发生这样的问题的部分基础。

但是,我们暂时往回一步,思考一下每个结果都有对应的发生概率这样的想法本身,特别是将这种想法应用到样本和总体。

统计学的许多推论方式是基于正态分布的,在本章我们已经对正态分布进行了一定程度的讨论。到现在为止都没问题。我们知道在理论和实践方面,人类(以及其他动物、物理、自然科学领域的现象)的许多特征的分布都是正态的(例如身高、体重以及智力水平),但是总体可能不是正态的。因此,当我们进行的许多研究是基于正态分布但是总体事实上不是正态的情况下,我们该怎么办?

别发愁。有一个非常重要也非常酷的定理——**中心极限定理**。就如人们所说,这个世界在一定程度上就是随机事件(也就是指随机数值),这个理论解释了某个正态分布的样本值的发生概率(也形成我们所用的推论工具的基础,这会在本书接下来和之后的章节中介绍)。

中心极限定理有两项基本的要素。

第一,与大量独立观察值相关的数值(如总和或均值)其分布近似于一个正态分布。第二,随着观察值或者样本量的增加这个分布会越来越趋向于正态分布。下面是非常有解释力的案例。

我们假定(从一个非正态分布中)得到 1 000 个数值大小从 1 到 10 的数据总体。你可以看到这样的总体分布可能类似于图 8.3 所示。不是很好看,也确实是没有可辨别类型的随机分布。

现在我们选择 100 个规模为 5 的样本，接着计算每一个样本的平均值。我们会得到 100 个平均值，如图 8.4 绘制图形时可以看到类似于某种正态分布。

这样的观察是将从样本获得的结果和将这些结果一般化到总体连接起来的关键。关键的假设是从总体中重复的抽样（即使总体分布是看起来很怪的正态分布或者明显就不是正态分布）会形成一个数据集，这个数据集接近正态分布。如果不是这样的情况，那么推论统计的许多参数检验（假定了是一个正态分布）就不能应用。

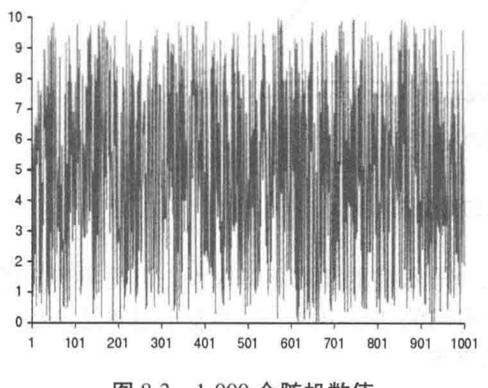


图 8.3　1 000 个随机数值

图 8.4　以图 8.3 所示数值为总体获取的 100 个平均数

更正态的曲线 101

你已经知道有三个主要的特征使得曲线成为正态曲线，或者看起来更像钟形，但是正态曲线不仅仅是这三个特点。认真地看图 8.5 中的曲线。

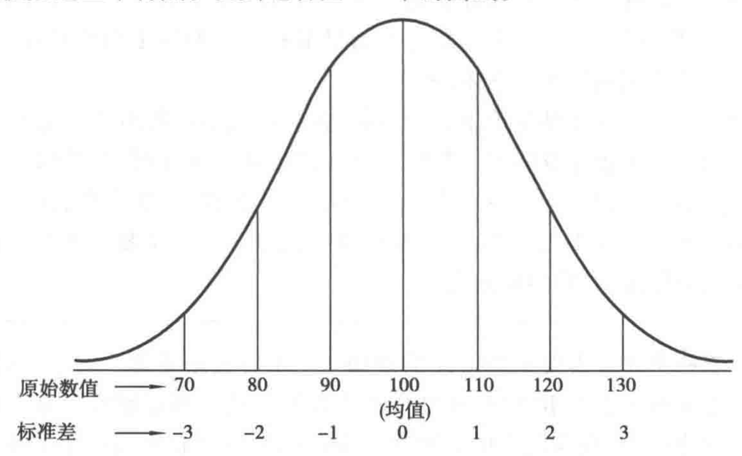


图 8.5　分成多个部分的正态曲线

这个分布的均值是 100，标准差是 10。我们已经在 x 轴上增加了表示数据分布中偏离于均值的以标准差表示的距离。你可以看到 x 轴（表示分布中的数值）的刻度是以 10（也就是分布的标准差）为间隔从 70 增加到 130，10 是一个标准差的值。我们编制了这些数字（100 和 10），所以不要去猜测我们如何得到的这些数字。

如此，短暂的回顾告诉我们分布的均值是 100，标准差是 10。曲线内的每条垂直线将曲线分成一个部分，每个部分由特定的值限定。例如，均值 100 右端的第一部分由数值

100 和 110 限定，这表示偏离均值（均值是 100）一个标准差。

而且在每一个原始数值（70，80，90，100，110，120，130）下端，你会发现相应的标准差（-3，-2，-1，0，+1，+2，+3）。就如你已经知道的，我们的案例中每一个标准差是 10。因此偏离均值（均值是 100）一个标准差就是均值加 10 或者是 110。不是很难，对吧？

如果我们深入进行讨论，你就能够看到均值为 100，标准差为 10 的正态分布表示的数值范围是 70 到 130（包括-3 到+3 个标准差）。

有关正态分布、均值和标准差的一个重要事实始终是正确的：对任何数值分布来说（不论均值和标准差的具体数值），如果数值是正态分布的，几乎 100%的数值处于均值的-3 到+3 个标准差范围内。这非常重要，因为这个事实适合所有的正态分布。这个规则的确很实用（再说一次，不论均值和标准差的具体数值），基于此数据分布之间可以相互比较。接下来我们还会讨论。

接着上面的内容，我们再进一步深入讨论。如果数值分布是正态的，我们也可以说一定百分比的数值会落在 x 轴的不同数据点之间（例如均值和一个标准差之间）。实际上数值分布中大约 34%（实际上是 34.13%）的数值落在了均值（在这个案例中是 100）和均值以上 1 个标准差（就是 110）的范围内。你可以永远确信这个事实。

想了解更多吗？认真地看图 8.6。你可以看到特征明显的同样的正态曲线（均值等于 100，标准差等于 10）——以及均值和标准差限定范围内我们预期出现的数值的百分比。

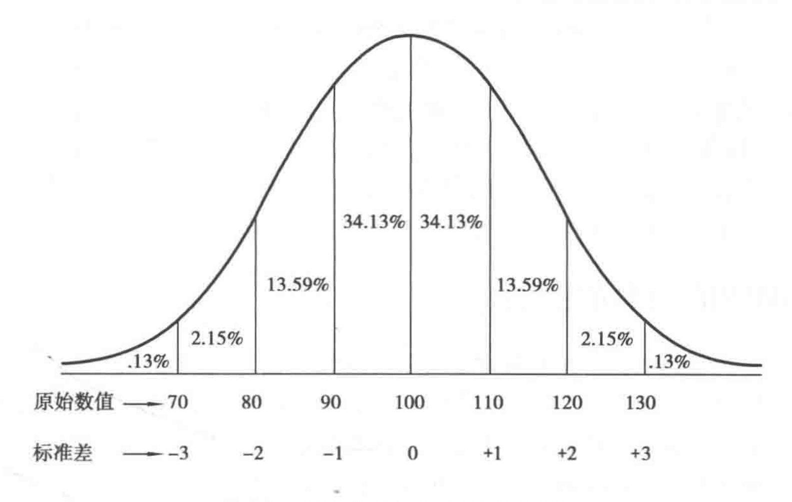


图 8.6 正态曲线下数据值的分布

下面是我们可以得出的结论。

两者的距离	包括的数据值	包括的数值范围 （如果均值=100，标准差=10）
在均值和 1 个标准差之间	曲线覆盖 34.13%的数据值	100—110
在 1 个标准差和 2 个标准差之间	曲线覆盖 13.59%的数据值	110—120
在 2 个标准差和 3 个标准差之间	曲线覆盖 2.15%的数据值	120—130
在 3 个标准差及以上	曲线覆盖 0.13%的数据值	130 以上

如果你将正态曲线每一半的百分值加起来，猜猜你会得到什么？很对，是 50%。为什么？正态曲线之下，均值和均值右侧所有数值的距离范围内包括了 50%的数值。

因为曲线是中心轴对称的(每一半是另一半的镜像),两个部分加起来表示100%的数值。虽然这并不高深,但是指出这一点很重要。

现在我们将这个逻辑应用到均值100左侧的数值。

两者间的距离	包括的数据值	包括的数值范围(如果均值=100,标准差=10)
在均值和1个标准差之间	曲线覆盖34.13%的数据值	90—100
在-1个标准差和-2个标准差之间	曲线覆盖13.59%的数据值	80—90
在-2个标准差和-3个标准差之间	曲线覆盖2.15%的数据值	70—80
在-3个标准差及以下	曲线覆盖0.13%的数据值	70以下

现在要记住的是我们使用的均值100和标准差10仅仅是特定案例的样本统计值。很明显,不是所有的分布都是均值为100、标准差为10。

所有的这些都有规律,特别是你会发现34.14%、13.59%这些数值独立于实际的均值和标准差。这些数字是源于曲线的形状而不是因为具体的均值和标准差。实际上,如果你在一个硬纸板上绘制正态曲线,接着将均值和一个标准差范围内的区域切下来,然后称出重量,重量恰好是从中切除曲线的整个硬纸板的34.13%。(试一次,这是真的。)

在我们的案例中,这意味着(粗略地说)68%(双倍的34.13%)的数值落在原始数据90到110之间。那么其他32%呢? 这个问题很好。一半(16%,或者13.59%+2.15%+0.13%)落在均值的一个标准差以上(均值右侧),另一半落在均值的一个标准差以下(均值左侧)。而且因为曲线向下倾斜,所以数值越偏离均值,曲线覆盖的区域的范围就越小,那么一个数值落在数据分布极值范围内的可能性要小于落在中间的可能性,这一点不惊奇。这也是为什么曲线在中间有波峰而没有任何方向的偏度。

我们最中意的标准值:z值

你已经多次看到数据分布在集中趋势和变异性方面如何的不同。

在一般的研究实践中,我们会发现所处理的分布相当不同,但是我们需要对它们进行相互比较。而进行这样的比较我们需要一定的标准。

这就是**标准值**(standard scores)。这些值以标准差为单位进行了标准化,所以是可比较的。例如,均值为50、标准差为10的分布的标准值1与均值为100、标准差为5的分布的标准值1是相同的。他们都表示一个标准值,并且与各自均值的距离也相同。我们也可以使用正态曲线的知识,并确定偏离均值一个标准差的数值出现的概率,会在后面继续探讨。

虽然还有其他类型的标准值,但是学习统计学过程中最常看到的是z值(z score)。z值就是原始数据与数据分布均值的差除以标准差所得的结果(见公式8.1)。

$$z=\frac{X-\overline{X}}{s} \tag{8.1}$$

其中:z是z值,

X是具体的数值,

$\overline{X}$是数据分布的均值,

s 是数据分布的标准差。

例如,应用公式 8.1 你可以看到如何计算均值是 100 原始数值是 110 标准差是 10 的 z 值。

$$z=\frac{110-100}{10}=+1.0 \tag{8.2}$$



在给定 z 值和其他数值的情况下很容易计算原始数值。你已经知道了给定原始数值、均值和标准差的 z 值公式。如果知道 z 值、均值和标准差,如何计算相应的原始数值?使用公式 $X=z(s)+\overline{X}$ 就可以计算。如果需要可以很容易地将原始数值转换为 z 值或者进行相反的转换。例如,数据分布的均值是 50,标准差是 5,标准值是 −0.5,那么原始数值 $X=(-0.5)(5)+50$,或 47.5。

下面是均值为 12、标准差为 2、样本规模为 10 的原始数值和相应的 z 值。均值以上的原始数值对应的 z 值是正数,反之,均值以下的原始数值对应的 z 值是负数。例如,原始数值 15 对应的 z 值是+1.5,原始数值 8 对应的 z 值是−2。当然与均值相等的原始数值 12(或均值)的 z 值是 0,因为原始数值与均值的距离为 0。

X	$X-\overline{X}$	z 值
12	0	0.0
15	3	1.5
11	−1	−0.5
13	1	0.5
8	−4	−2.0
14	2	1.0
12	0	0.0
13	1	0.5
12	0	0.0
10	−2	−1.0

依据对以上数值的观察进行简要的总结。

首先,均值以下的数值(如 8 和 10)对应的 z 值是负数,而均值以上的数值(如 13 和 14)对应的 z 值是正数。

其次,正的 z 值一般落在均值的右侧,也就是数据分布的上半部分。负的 z 值通常落在均值的左侧,也就是数据分布的下半部分。

再次,我们讨论落在均值一个标准差以上的一个数值时,也就是说这个数值是在均值的一个标准值之上。就我们的目的来说,比较数据分布的数值时使用标准差和 z 值是相当的。换句话说,z 值就是偏离均值的标准差的个数。

最后也是最重要的一点,不同分布的 z 值具有可比性。我们应用下面的数据表说明这一点,这个数据表和上一个类似。我们从 100 个数值中选了 10 个样本数值,样本分布的均值是 58,标准差是 15.3。

原始分数	$X-\overline{X}$	z 值
67	9	0.59
54	−4	−0.26
68	10	0.65
33	−25	−1.63
56	−2	−0.13
76	18	1.18
65	7	0.46
35	−23	−1.50
48	−10	−0.65
76	18	1.18

在之前看到的均值是 12、标准差是 2 的数据分布中，原始数值 12.8 对应的 z 值是 +0.4，也就是说原始数据 12.8 距离均值 0.4 个标准差。在均值是 58 标准差是 15.3 的第二个数据分布中，原始数值 64.2 对应的 Z 值也是+0.4。这是奇迹吗？不，这只是一个很好的想法。

原始数值 12.8 和 64.2 相对来说与均值的距离相等。如果这些原始数值用标准值表示，就可以直接使用在各自分布中的相对位置来直接进行相互比较。

使用 Excel 计算 z 值

使用 Excel 计算 z 值(或者标准值)不难，只需要使用一个简单的公式。在图 8.7 中，你可以看到所用的是之前的一组数据，均值是 12，标准差是 2。在公式栏是计算 z 值所用的公式。这个公式可以从单元格 B2 复制到单元格 B11，非常容易和快捷。

B3 =(A3-(AVERAGE($A$2:$A$11)))/STDEV($A$2:$A$11)

	A	B
1	X	z score
2	12	0
3	15	1.5
4	11	-0.5
5	13	0.5
6	8	-2
7	14	1
8	12	0
9	13	0.5
10	12	0
11	10	-1

图 8.7 使用 Excel 计算 z 值

我们这里学习了一些新的内容。

首先，我们在公式中使用了函数。我们使用函数计算均值和标准差，接着使用一个公式计算 z 值，我们跳过了其中使用函数的步骤说明。计算中，我们没有直接使用下面的 z 值公式。

$$z = \frac{X - \overline{X}}{s} \tag{8.3}$$

我们使用的公式形式如下：

= (A3-(AVERAGE($A $2: $A $11)))/STDEV($A $2: $A $11)

在这里，与均值的偏差通过原始数据（在单元格 A3）减去均值（使用 AVERAGE 函数）得到。所得结果接着除以标准差，标准差使用 STDEV 函数计算。使用第二个公式的好处是可以节省建立不同的值并将不同的值代入公式的步骤。

其次，所有美元符号 $ $ $ $ 的作用是什么？很简单，就是指那些单元格的绝对位置而不是相对位置——我们所指的单元格处在特定的位置，它们不会因为计算结果所在单元格位置的变化而发生改变。我们输入的单元格地址是绝对的位置，因为这些单元格是用于计算平均值和标准差，它们不会发生改变，而在其中计算 z 值的单元格可以发生变化。这就是为什么在图 8.7 中你可以看到单元格 A3 没有美元符号。这是相对的单元格位置，会随着公式沿着列向下复制时而发生变化。

更多 Excel

Excel 有很好的函数（STANDARDIZE 函数）为任何原始数据（就如我们所展示的数据，任何的原始数据集）来计算标准值。函数的形式如下。

=STANDARDIZE (X, mean, standard deviation)

其中：X 是你想要计算 z 值的原始数值，

mean 是数据集的均值，

standard deviation 是数据集的标准差。

例如，如果你想计算均值为 95、标准差是 2.3 的数据集中的一个原始数据 98 的 z 值，函数形式如下。

=STANDARDIZE(98,95,2.3)

见图 8.8 所示。在单元格 B1 是原始数据，单元格 B2 是均值，单元格 B3 是标准差，单元格 B4 是计算 z 值的函数，z 值是 1.30。

B4 =STANDARDIZE(B1,B2,B3)

	A	B	C	D
1	Raw Score	98		
2	Mean	95		
3	s	2.3		
4	z score for a raw score of 98	1.30		

图 8.8 使用 STANDARDIZE 函数计算 z 值

计算一个 z 值的工作量会很大。因此，我们将这个函数应用于计算一组 z 值。如图 8.9 所示。看起来是否很熟悉？工作表中的数据也是我们计算 z 值的第一个案例中所用的数据（见图 8.7）。但是我们使用 STANDARDIZE 函数计算一组数据的 z 值，与 AVERAGE 函数和 STDEV 类似，将公式在对应的列中向下复制。下面是我们公式中所用的 STANDARDIZE、AVERAGE 和 STDEV 函数。

=STANDARDIZE(A2, AVERAGE($A$2: $A $11), STDEV($A $2: $A $11))

如图 8.9 所示，公式发挥了作用。

B3 =STANDARDIZE(A3,AVERAGE($A$2:$A$11),STDEV($A$2:$A$11))

	A	B	C	D	E	F	G	H	I	J
1	X	z score								
2	12	0								
3	15	1.5								
4	11	-0.5								
5	13	0.5								
6	8	-2								
7	14	1								
8	12	0								
9	13	0.5								
10	12	0								
11	10	-1								

图 8.9　在公式中使用 STANDARDIZE 函数计算 z 值

z 值表示什么

你已经了解特定的 *z* 值表示一个原始数值，也表示数据分布在 x 轴上的特定位置。而且 *z* 值越大（如−2 或+2.6），距离均值越远。

因为你已经知道落在 x 轴的特定两点之间的面积占全部曲线覆盖面积的百分比，例如，均值和+1 个标准差之间的面积是 34%，+1 个标准差和+2 个标准差之间的面积是 14%，我们也可以得出如下正确的结论：

- 84%的数值落在值为+1 的 *z* 值之下（50%落在均值之下，34%落在均值和值为+1 的 *z* 值之间）。
- 16%的数值落在值为+1 的 *z* 值之上（曲线下的全部面积是 100%，84%的数值落在值为+1.0 的 *z* 值之下）

想一下上面的两个结论。我们所要表达的是，如果数据分布是正态分布，曲线的不同面积可以用标准差或者 *z* 值的不同数值来表示。当然，你应该明白这些面积或比例也可以被看作是表示特定数值出现的概率。例如，这里有一个问题：

> 在均值为 100 标准差为 10 的数据分布中，110 或 110 以上的数值出现的概率是多少？

答案是 16%，或者说 100 次中会出现 16 次。我们如何得到答案？

首先我们计算对应的 *z* 值，也就是+1〔(110−100)/10〕。接着，依据我们已了解的知识（见图 8.6），我们知道 *z* 值为 1 表示 x 轴上的一个特定位置，数据分布中 84%（50%加 34%）的数值落在这个位置之下，这个位置之上是 16%的数值或概率是.16。因为我们已经知道均值和均值之上或之下 1、2 和 3 个标准差之间的面积，我们可以很容易地得出任何 *z* 值对应的数值出现的概率。

但是我们介绍的方法对于 *z* 值是 1、2 和 3 很合适。但是如果 *z* 值不是像 2 一样的整数，而是像 1.23，−2.01 一样的非整数，该如何确定呢？我们需要找到更简单的方法。

我们怎么办呢？很简单，就是学会微积分并将微积分应用到正态曲线来计算 x 轴上每一个可能的值对应的曲线面积，或者我们使用附录 B 的 B1 表（正态分布表）。表中列出了不同 *z* 值对应的曲线之下的所有面积数值，当然不包括极值。这个表有两列。第一列是 *z* 值，就是已经计算的 *z* 值。第二列是均值和 *z* 值之间的面积，就是这两点之间曲线所覆盖的面积。

例如，如果我们想知道均值和值为+1 的 *z* 值之间的面积，在 *z* 值列找到数值 1.00，接着在对应的第二列找到均值和 *z* 值 1.00 之间的面积数值是 34.13。你以前看到过这样的

表吗?

为什么在表中没有正号或负号,如-1.00 呢? 因为曲线是对称的,z 值是正是负对数值没有影响。均值和 1 个标准差之间的面积在任何方向上都是 34.13%。

接着往下。对于特定的 z 值如 1.38 来说,你想知道与 z 值对应的概率。如果你想知道均值和 z 值 1.38 之间的面积百分比,你可以在 B1 表中找到对应 z 值 1.38 的面积是41.62,这表明数据分布中41%以上的数据落在 z 值0 和 1.38 之间,而 92%(50%+41.62%)的数据落在 z 值 1.38(包括 1.38)之下。现在你肯定已经注意到最后的案例中根本没有提到原始数值。一旦你会用这个表,就不再需要原始数值。

但是我们是否始终只对均值和某些 z 值之间的面积感兴趣? 要不要关注两个 z 值之间的面积呢? 例如,我们有兴趣知道的是 z 值 1.5 和 z 值 2.5 之间的面积,或者说数值落在这两个 z 值之间的概率是多少? 我们如何使用这个表计算这些结果? 这很容易。只要找到每一个 z 值对应的面积,然后用一个减去另一个。通常绘制如图 8.10 所示的图有助于我们理解。

例如,我们想找到均值为 100,标准差为 10 的数据分布中原始数值 110 和 125 之间的面积。我们可以采用如下的步骤。



1—计算原始数值 110 对应的 z 值,也就是(110-100)/10,即+1。
2—计算原始数值 125 对应的 z 值,也就是(125-100)/10,即+2.5。
3—使用附录 B 的 B1 表,找到均值和 z 值+1 之间的面积,是 34.13%。
4—使用附录 B 的 B1 表,找到均值和 z 值+2.5 之间的面积,是49.38%。
5—因为你想知道两个 z 值之间的距离,用较大的数减去较小的数,也就是 49.38%-34.13%,即 15.25%。图 8.10 中包含很多信息。

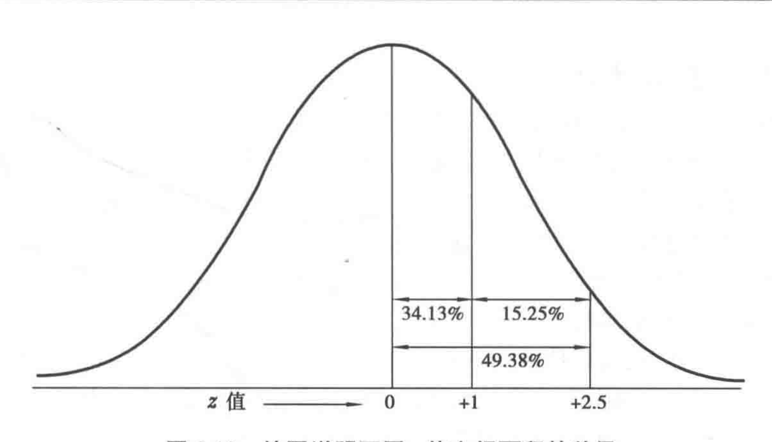


图 8.10 绘图说明不同 z 值之间面积的差异

更多 Excel

这里还有另外的方法计算特定原始数值的概率,可以使用 NORM.S.DIST 函数。这个函数的形式如下:

=NORM.S.DIST(z,cumulative)

其中：

z 是你想确定其概率的原始数值对应的 z 值，

cumulative 是逻辑值。[在我们的案例中几乎都是真实的(true)]

因此，你想知道值为 1 的 z 值的概率，那么函数形式如下：

=NORM.S.DIST(1, true)

最后结果就回到单元格，值为 0.841345，这个值看起来很熟悉吧。试着使用一下这个函数。这是非常好用的方法，如果没有附录 B1 表，就可以使用这个函数。

因此我们十分确信，特定数值出现的概率可以通过确定这个数值相对其他数值落在数据分布的某个区域来理解。在这个案例中，数值出现在 z 值+1 和 z 值+2.5 之间的概率大约是 15%。

这儿还有另一个例子。在均值 100，标准差 10 的数据集中，原始数值 117 对应的 z 值是 1.70。这个 z 值对应的曲线覆盖面积是 95.54%(50%+45.54%)，意味着数值出现在 z 值 1.70之下的概率是 95.54%，或者说 100 个中有 95.5 个，或者是 0.955。



关于标准值的两个方面

首先，虽然我们关注的重点是 z 值，但还有其他类型的标准值。例如，T 值是另一种类型的标准值，通过 z 值乘以 10 加 50 来计算。这个标准值的优点是很少出现负值。和 z 值一样，T 值也可以比较不同数据分布的数值。

其次，标准值和标准差完全不同。标准值来自预先确定均值和标准差的数据分布。考试成绩如 SATs 和 GREs(研究生入学考试)使用标准化分数，可以很容易地比较均值和标准差相同的数据分布的数值。

z 值真正表示什么

统计游戏的一个作用是能够估计某个结果的概率。如果我们掌握了前面的内容并且能够在本章学习中再往前走一步，就能确定某些特定事件发生的概率。接下来我们使用一些标准来判断我们考虑的事件的发生是高于还是低于我们预期的随机概率水平。研究假设提出了预期事件发生的命题，我们使用统计工具来估计事件发生的概率。

这是"统计学是什么"的"20 秒版本"，但是包含了很多内容。现在我们回顾这一段所有的内容，并通过案例来理解。

比如说你的老朋友，值得信赖的卢给了你一枚硬币，并让你决定硬币是否是"真的"。如果是真的，那你抛十次硬币，可以得到 5 次正面 5 次反面。因为每抛掷一次出现正面或反面的概率是 0.5，所以我们预期出现 5 次正面和 5 次反面。在 10 次独立的抛硬币实验中(也就是一次抛掷不会影响下一次)，我们应该得到 5 次正面，等等。现在的问题是出现多少次正面会确认这枚硬币是伪造的或者非法的?

现在说说我们用以判断"真"的标准，也就是如果抛掷 10 次硬币我们得到正面(或者反面)的次数少于 5%次，我们就能说硬币是伪造的，那就应该叫警察来抓卢了(或者他本来就是在假释中)。5%就是统计学家们使用的标准。如果事件(正面出现的次数、一次考试的成绩或者两个数据组平均值的差异)发生的概率是极值(我们所说的极值定义为出现

次数小于 5%），那么就是不可能出现的结果，在这个案例中就是“硬币不是真的”。

这是 10 次随机抛掷硬币实验中预期出现的正面次数的分布。所有可能的结果组合是 2^{10}或 1 024 种，例如 9 次正面 1 次反面，7 次正面 3 次反面，10 次正面 0 次反面，等等。而 10 次抛掷实验中得到 6 次正面的概率大约是 21%。

正面次数	概　率
0	0.00
1	0.01
2	0.04
3	0.12
4	0.21
5	0.25
6	0.21
7	0.12
8	0.04
9	0.01
10	0.00

以上是任何特定结果出现的可能性，如 10 次投掷中出现 6 次正面的可能性大约是 .21，或 21%。现在是做出决定的时刻。也就是 10 次抛掷中你得到多少次正面就可以断定硬币损坏了、不均匀，或者是伪造的。

与优秀的统计学家一样，我们采用的标准也是 5%，之前我们也是这样定义。如果观察到的结果（我们投掷硬币的结果）的概率小于 5%，我们可以断定这是不可能的，除非有非概率事件出现——我们的结论就是“非概率事件”是伪造的硬币。

看看上面的数据表，你就会发现 8、9 或 10 次正面出现的结果都小于 5%。因此，如果 10 次硬币投掷的结果是 8、9 或 10 次正面，结论就是这枚硬币不是真的。（当然正面出现 0、1 或 2 次也能得出相同的结论，这是对的。硬币另一面的出现情形也如此。）

相同的逻辑也适用于在之前关于 z 值的讨论。我们预期 z 值多大就可以断定一个结果的出现不是由于随机因素而是由于一些其他因素？如果查阅附录 B 的正态曲线表你会知道 z 值的临界点 1.65 包含了曲线覆盖面积的 45%，如果与曲线另一侧覆盖面积的 50% 加起来，就得到全部面积的 95%。也就是在 x 轴的这一点之上留下了 5%。任何表示 z 值 1.65 或这个值以上的数值就在很小的区域内——或者至少是分布在出现的概率小于其他数值的区域内。

假设检验和 z 值：第一步

我们所要说的是任何事件都有相应的发生概率。我们使用这些概率值来推断我们预期的某个事件不出现的可能性。例如，抛掷 10 次硬币出现 1 次正面 9 次反面的可能性就很小。我们也说过如果事件的发生 100 次中只有 5 次（5%），我们就可以认为相对于其他可能发生的事件来说这个事件更不可能发生。

研究假设相关的结论当然也同样如此。零假设（见第 7 章）认为群体或变量之间没有差异，而且发生的可能性是 100%。我们要尽力检验零假设可能存在的错误。



相对于 t 值和 t 检验来说，z 值非常好用，检验相关内容会在第 11 章和 12 章学习。现在，对你来说重要的是知道 z 检验常用于总体，而 t 检验常用于样本。这是为什么在期刊论文和研究报告中你经常看到 t 检验而不是 z 检验。

换句话说，如果通过研究假设的检验我们发现事件发生的可能性是极值，那么研究假设就是更有力的解释而不是零假设。因此，如果我们发现 z 值是极值（发生概率小于 5% 就是极值），我们就会说极值出现的原因不是由于随机因素而是与某种关系或者某种处理方式有关。我们会在接下来的章节更详细地讨论这一点。

小　结

理解整个推论统计的最基本的和最重要的技能就是可以计算 z 值，并能够估计在一个样本数据中某个数值出现的可能性。一旦我们知道考试成绩或者组间差异发生的可能性，我们就能将这个可能性和我们随机预期的可能性进行比较，并进一步得出推论结论。在《爱上统计学：Excel》第Ⅳ部分的开始，我们将这个模型应用到检验差异性问题的特定案例中。

练　习

1. 回答下面有关正态曲线的问题。
   a. 正态曲线有什么特点？
   b. 你认为人类的哪些行为、特征呈正态分布？
2. 为什么 z 值是标准化的值？为什么 z 值可以用于比较不同数据分布的数值？
3. 应用 z 值的主要的价值是什么？提供一个案例说明如何使用 z 值。
4. 计算下面的均值为 50、标准差为 5 的数据分布的原始数值的 z 值。
   a. 55　　b. 50　　c. 60　　d. 58.5　　e. 46
5. 假定 z 值是 1.5，均值是 40、标准差是 5，相应的原始数值是？相应的 t 值是？
6. 问题 6a 到 6b 依据均值为 75、标准差为 6.38 的数据分布来回答。绘制一个简图来确定你需要怎么做。
   a. 一个数值落在原始数值 70 和 80 之间的概率是多少？
   b. 一个数值落在原始数值 80 以上的概率是多少？
   c. 一个数值落在原始数值 81 和 83 之间的概率是多少？
   d. 一个数值落在原始数值 63 以下的概率是多少？
7. 使用 NORM.S.DIST 函数计算一个数值落在 z 值+1 和+2 之间的概率。要记住，你不需要任何原始数值，因为 z 值为 1。
8. 这个练习使用第 8 章数据集 1。使用 NORM.S.DIST 函数计算第一个原始数值也就是单元格 A2 的数值发生的概率，接着复制函数计算每个数值相应的 z 值的概率。（提示：在函数中使用单元格引用，而不是实际的单元格数值。）
9. Jake 为了得到身体健康的证明需要取得组内前 10% 的成绩。班级的均值是 78，标准差是 5.5，Jake 需要得多少分才能得到他需要的证明？

10.理解中心极限定理的主要价值是什么？

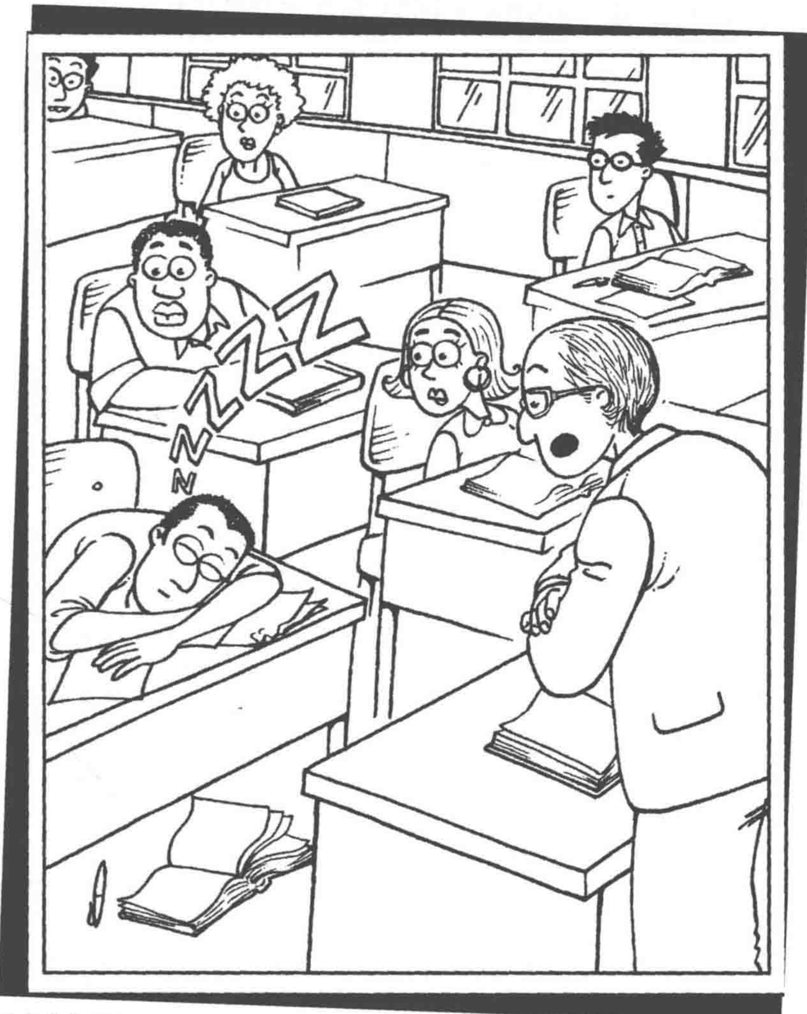
11.为什么简单的组合没有意义？例如，检验不同课程的考试得分——只计算平均值，然后就这样结束？

12.相对于他或她的同学，谁的学习更好？下面是你可能需要的所有信息？

数　学			
班级均值	81		
班级标准差	2		
阅　读			
班级均值	87		
班级标准差	10		
原始值			
	数学成绩	阅读成绩	平均值
Noah	85	88	86.5
Talya	87	81	84
z 值			
	数学成绩	阅读成绩	平均值
Noah			
Talya			

第Ⅳ部分

显著性差异——使用推论统计



“我看大卫这会儿正忙着计算z值……”

你已经学到这里，而且仍然充满活力，所以要恭喜你。到现在为止，你已经很好地掌握了描述统计是什么，偶然性因素在决策中如何影响结果，以及由于偶然性因素和处理方法的不同，结果出现的可能性是如何的。

你是建立和理解假设在社会和行为科学研究中所扮演的角色的专家。现在就是实践的时候了。让我们看看在《爱上统计学：Excel》接下来的部分会学到什么。最重要的是你曾经付出的努力会以理解具体的问题而得到快速的补偿。

这部分的主要内容是理解和应用特定的统计类型来回答特定类型的研究问题。我们会介绍最常用的统计检验，也会介绍一些稍微复杂的统计检验类型。在最后部分会介绍一些比较常用的统计软件包，这些统计软件包可用于计算那些我们用老式的计算器计算的相同的值。

我们以显著性概念的简短讨论开始，然后逐步进行推论检验。接着就是特定检验的案例。这一章需要动手的内容挺多，现在就开始吧。

显著性的显著——对你我来说意味着什么 9

本章你会学到什么

- 显著性的概念和显著性的重要性
- 第一类错误和第二类错误的重要性和两者之间的区别
- 如何进行推论统计
- 如何根据目的选择适当的统计检验

显著性的概念

对于初学统计学的学生来说可能没有哪一个概念或词汇比统计显著性更让人迷惑了。但是,这并不意味着事实就是如此。虽然统计显著性是非常重要的概念,但它也很简单,基础统计学课上的任何学生都可以理解。

我们需要用研究案例来说明我们总结的要点。例如达科特和理查德(E.Duckett & M.Richard)的"单亲母亲家庭中母亲的职业和青少年的日常经验(Maternal Employment and Young Adolescents' Daily Experiences in Single Mother Families.)"(1989 年密苏里州堪萨斯,儿童发展研究协会论文)。他们调查了 436 个 5 年级到 9 年级的青少年对母亲就业的态度。

特别的是他们调查了母亲有工作和没工作的青少年在态度上是否存在差异。他们也检验了其他因素,但是就这个案例来说,我们关注的主要是有工作的母亲和没有工作的母亲之间的群体差异。再加一项内容就是在是否存在差异的讨论中增加显著性,因此我们的研究假设类似如下的陈述:

> 依据情绪状态的测量,母亲工作和不工作的青少年对母亲就业的态度具有显著性差异。

我们所说的**显著性**的含义是指两个群体的态度之间的任何差异是由于系统性因素而不是偶然性因素的影响。在这个案例中,影响因素是母亲是否工作。我们假定控制了可能影响两个群体之间差异的所有其他因素。因此,余下的解释青少年态度差异的唯一因素就是母亲是否工作。这对吗? 是的。这就完了? 不没呢。

是否只有我们是完美的

我们的世界并不完美,所以我们在一定程度上在我们确定引起群体间差异的那些因

素时要留有余地。换句话说，你需要表明即使你很确信两个青少年群体之间的差异是由于母亲的就业状态引起的，但是你不能 100%、绝对地、肯定地、不置疑或毫不含糊地确信这一点。你的结论是错误的可能性始终存在，不论这个可能性多小。

为什么？原因很多。例如，有可能你的结论完全错误。也许在这次研究中，青少年的态度差异不是由于母亲工作或不工作引起的，而是由于其他没有注意到的解释因素，例如当地在职母亲俱乐部举办的演讲，一些学生加入了这样的俱乐部。如果一个青少年群体的成员几乎都是男性，而另一个青少年群体的成员几乎都是女性，这样的情况下又该是怎样的结论？这也可能是差异的来源。如果你是优秀的研究者，也进行了一定的研究，你可以解释差异，但是总有可能你做不出解释。你必须考虑这种可能性的存在。

那么你要怎么做？在大多数涉及假设检验（例如这里的案例中的群体差异）的科学研究中一定存在一定数量的不能控制的误差——这也是前面几章已经讨论过的偶然性因素。你愿意承担的风险水平或者概率水平就是**显著水平**，这个词不会给人们带来内心的恐惧。

显著水平（significance level，这里是快捷简单的定义）是指不能 100%确信实验中观察到的结果是由于处理因素或需要检验的因素引起的所要承担的风险，这个因素在我们的案例中是母亲是否工作。如果你看到的陈述是显著性结论在 0.05 概率水平下（或者更专业地说 $p<0.05$，在专业期刊中你会经常看到），换句话说就是 20 次中有 1 次（或 0.05，或 5%）所发现的任何差异不是由于假定的原因（母亲是否工作），而是由于其他未知原因引起的。你的工作就是尽可能减少这种可能性，消除所有可能引起观察到的差异的任何其他原因。因为你不可能完全消除这种可能性（没有人能够控制所有潜在的因素），必须分配一定的概率水平，并谨慎地陈述结论。

总之（实际上也是如此），研究者定义了他或者她愿意承担的风险水平。如果结果落在这个范围内，也就是说“这不是偶然出现的——而是有其他因素在产生影响”，研究者就会知道零假设（表达式是等式）不是观察到的结果的最有力的解释。相反研究假设（也就是说是不等式，或者说存在差异）是可选择的解释。

现在来看另一个案例，这是假设的案例。研究者有兴趣了解参与学前项目的孩子与没有参与的孩子在学习成绩上是否存在差异。零假设是两个群体的学习成绩相等。研究假设是参与学前项目的孩子的平均成绩高于没有参与项目的孩子的平均成绩。

作为一个优秀的研究者，你的工作就是表明（尽你所能就好，没有人可以完美地解释任何事物）两个群体之间存在的任何差异仅仅是由于学前项目的影响，而不是任何其他因素或者因素的组合。你可以使用一些技术（你会在高级的统计学课程中学习）控制或者消除影响差异的所有可能的原因，如父母教育水平、家中孩子的数量等的影响。一旦消除了其他潜在的解释变量，唯一留下的对差异的解释就是学前经验的影响。

但是你可以绝对地确定吗？不，你不可以。为什么？首先，你不能确定你所研究的样本能够很好地代表总体。而且即使样本能很好地代表总体，也总是存在影响结论的影响因素，而且在设计试验的过程中你总会无意中遗漏这些因素。研究中始终存在犯错的可能性。

如果推断考试成绩的差异是由于经历的不同，就得承担一定的风险。实际上（给点掌声鼓励）风险水平就是你愿意执行的统计显著性水平。

统计显著性（statistical significance，这里是正式的定义）是指零假设为真的情况下拒绝

零假设所要承担的风险水平。就上面的例子来说,零假设是两个样本群体之间没有差异(记住,零假设始终以等式的形式表述)。但是在给你们的数据中,你们会发现差异确实存在。也就是考虑到目前你们找到的证据的情况下,群体成员的身份似乎对学习成绩有影响。但是在真实的世界中却可能没有差异。如果你拒绝了你陈述的零假设,就犯了一个错误。犯此类错误你可能承担的风险(或者说显著水平)就是人们熟知的第一类错误。

世界上最重要的表格(只对这一学期而言)

下面是简要的概括。

零假设可能是真实的也可能是虚假的。两个群体之间可能真的没有差异,也可能真的确实是不相等的(如两个群体之间存在差异)。但是要记住你永远不会知道真实的状况,因为零假设不能直接检验(记住零假设只应用于总体)。

实际上,作为优秀的统计学者你既可以选择拒绝也可以接受零假设,对吧?一共有四种情况,你可以在表 9.1 中看到。

现在就来看看表中的每一格。

更多地了解表 9.1

表 9.1 中的四个重要的格描述了零假设的性质(真实的或虚假的)和相应的选择(接受或拒绝零假设)之间的关系。就如你所看到的,零假设可能是真实的也可能是虚假的,而你可能拒绝也可能接受它。

对于理解这个表来说最重要的事实是研究者永远不知道零假设的真实性质,以及群体之间真的存在还是不存在差异。为什么?因为总体(零假设所表示的)不能直接检验。为什么?这样做是不现实的,而这也是为什么我们有推论统计。

- 表 9.1 中格子 1 表示的状况是,零假设是真实的(群体之间没有差异)情况下研究者做出了接受的正确选择。这里没有问题。在我们的案例中,我们的结论表明两个孩子群体之间没有差异,而我们接受了零假设(也就是没有差异)的正确选择。
- 格子 2 表示的是严重的错误。这里我们拒绝了零假设(也就是没有差异),而零假设实际上是真实的(即真的没有差异)。两个孩子群体之间没有差异,但是我们得出的结论是有差异,这就犯一种类型的错误。这就是人们所说的**第一类错误**(Type Ⅰ error),也就是显著水平。
- 还存在一类错误。格子 3 里面我们已经接受了零假设(也就是没有差异),而零假设实际上是虚假的(即实际上存在差异)。两个孩子群体之间有差异,但是我们得出的结论是没有差异。很明显这就是人们熟知的**第二类错误**(Type Ⅱ error)。
- 表 9.1 中的格子 4 表示的状况是在零假设实际上是虚假的情况下,研究者做出了拒绝的正确决定。这里没有问题。在我们的案例中,我们的结论表明两个孩子群体之间有差异,而且我们做出了拒绝表述为没有差异的零假设的正确选择。

表 9.1 不同类型的错误

		可能的选择	
		接受零假设	拒绝零假设
零假设的真实性质	零假设是真实的	1 ☺对啦，零假设是真实的情况下你接受了零假设，而且群体之间没有差别。	2 哎！你犯了第一类错误，在群体之间没有差异的情况下拒绝了零假设。第一类错误也可以用希腊字母 α 表示。
	零假设是虚假的	3 哦！你犯了第二类错误，接受了虚假的零假设。第二类错误也可以用希腊字母 β 表示。	4 ☺很好，在群体之间存在差异的情况下你拒绝了零假设。也可以叫作检定力，或 $1-\beta$。



如果 0.05 是好的，那么 0.01 更好，为什么不把第一类错误的风险水平定在 0.000 001？对于每一个好的解释来说，有时你如此严谨地拒绝虚假的零假设以至于会错过了正确的零假设。这样严格的第一类错误的发生率留下的余地很小——实际上研究假设可能是正确的，但是相应的概率水平是 0.015——非常少见，但是由于严格的显著水平而错过了。

回顾第一类错误

我们现在把重点更多地放在格子 2 上，也就是犯了第一类错误，因为这是我们讨论的重点。

第一类错误或者说显著水平具有特定的值，而且与任何零假设检验中定义的你愿意承担的风险联系在一起。一般设置的显著水平是在 0.01 与0.05之间。

例如，如果显著水平是 0.01，这意味着在任何一个零假设检验中，只有 1%的可能性是零假设为真，而你拒绝了零假设，并且在群体之间实际上根本没有差异的情况下得出群体之间有差异的结论。

如果显著水平是 0.05，这意味着在任何一个零假设检验中，只有 5%的可能性是零假设为真，而你拒绝了零假设（并得出群体之间有差异的结论），而实际上根本没有群体间差异。要注意的是显著水平与零假设的独立性检验相关，而且不可以说“对零假设的 100 次检验中我犯了 5 次或总数中 5%的错误”。

在最近的研究报告中，统计显著性通常以 $p<0.05$ 表示，可以读作“观察到这种结果的概率小于 0.05”，在专业文章的报告中简单地表述为“在0.05的显著水平下”。



依据流行的统计分析软件的介绍，不再需要担心如“$p<0.05$”或“$p<0.01$”这样的陈述的不准确性——$p<0.05$ 就意味着是从 0.000 到 0.049 999 的所有值吗？但是类似 SPSS 的软件给出了你犯第一类错误愿意承担的风险的确定概率如 $p=0.013$，或者$p=0.158$。因此当你在研究报告中看到类似的陈述如“$p<0.05$”时，意味着 p 值是从 0.00 到0.049 999 999 999…之间

的任何值(你会明白的)。同样的,当你看到“$p>0.05$”或“$p=n.s.$”(非显著性),意味着拒绝真实的零假设的概率超过0.05,实际上可能性范围是0.050 000 1到1.00。

因此,知道一个结果的确定的概率是很大的进步,因为我们能够更准确地测量我们愿意承担的风险。

但是如果 p 值刚好等于0.05怎么办呢? 假定你已经了解了相关内容,如果你按照规则来解释,那么结果就是不显著的。一个结果要么是要么不是,因此0.049 999 999 99是显著的而0.05不是。如果Excel或者其他软件产生的 p 值是0.05,就扩展小数点后的位数——也许实际的结果是0.049 999 999 99。

在第一类错误之外,你还可能犯另一类型的错误,如在表9.1中所示。当你无意中接受了虚假的零假设就产生了第二类错误(表中的格子3)。



当我们讨论一项发现的显著性时,你可能听到检定力这个词。检定力是关于统计检验如何很好检验和拒绝一个虚假的零假设的概念。从数学上来说,检定力就是用1减去第二类错误的值。更有检定力的检验总是比没有检定力的检验更具有吸引力,因为有检定力的检验可以让你更接近辨别虚假的核心。

例如,样本群体代表的两个总体之间存在真实的差异,但是你错误地得出差异不存在的结论。

就理想状态而言,你想同时减少第一类错误和第二类错误,但是这总是很难实现,或者说很难控制。你已经完全控制了第一类错误的水平或者说你愿意承担一定的风险(因为你实际上设置了风险水平)。第二类错误没有直接控制,但是第二类错误与样本规模等因素相关。第二类错误对样本中个体的数量特别敏感,个体数量越多,第二类错误就越低。换句话说,样本的特征越是与总体的特征匹配(可以通过增大样本的规模提高),你接受虚假的零假设的可能性就越低。

显著性 VS.意义

对研究者来说,有趣的状况是发现试验的结果在统计上是显著的。你知道统计显著性的含义——也就是研究在统计上成功了,而且零假设不是发现结果的合理解释。如果你的实验设计和其他因素都经过谨慎的考虑,统计上显著的结论毋庸置疑使你对所在研究领域的贡献迈出了第一步。但是,统计显著性的值及其重要性或意义必须认真考虑。

例如,我们面对的情况是将规模非常大的不识字的成年人(大概10 000人)样本分成两组。一组通过使用电脑来接受集中的阅读训练,另一组通过课堂教学来接受集中的阅读训练。组1(在课堂中学习)的阅读测验的平均成绩,也就是被解释变量是75.6;组2(使用电脑学习)的平均成绩是75.7。两组的方差几乎相等。就如你所能看到的,成绩均值的差异只有十分之一点(75.6与75.7),但是对独立均值之间的显著性进行 t 检验时,结果在0.01的水平是显著的,这表明电脑学习者比教室教学学习者学得更好。(接下来的两章讨论 t 检验)

两组之间0.1的差异是统计显著的,但这是否有意义? 考试成绩(在这么小的范围内)的提高是否能够为花费300 000美元建立电脑辅助教学项目提供充分的合理解释? 或

者说这个差异无足轻重可以忽略，即使在统计上是显著的？

下面列出的是我们依据这个案例和其他许多可能的案例得出的有关统计显著性的重要性的结论。

- 统计显著性本身或内部是无意义的，除非所执行的研究具有合理的概念基础，可以由此推导结果具有显著性的意义。
- 统计显著性不能脱离发生的背景独立地解释。例如，如果你是学校系统的管理者，如果留级项目以半分之差显著地提高学生的标准化考试成绩，你是否愿意将学生留在一年级？
- 虽然统计显著性是很重要的概念，但不是终极目标，当然也不应该是统计研究的唯一目标。这就是为什么我们是检验假设而不是证明假设。如果我们的研究设计正确，那么甚至零假设也会揭示重要的信息。如果特定的处理因素没有产生影响，这也是其他人需要知道的重要信息。如果你的研究设计得很好，接着你应该知道为什么处理因素没有发挥作用，那么沿着这条线研究的其他人在设计自己的研究时就可以考虑你所提供的有价值的信息。

推论统计介绍

描述统计用于描述样本的特征，而推论统计是基于样本特征推断总体的某些特征。

在《爱上统计学：Excel》前半部分的几个方面，我们已经强调，好的科学研究的一个标志就是以这样的标准选择样本，即选出来的样本是总体的代表。接下来的过程就是推论，就是基于样本的检验（和试验）结果从较小的样本群体推断较大的群体。

在我们开始讨论具体的推论检验之前，先了解推论方法应用的逻辑。

推论如何进行

下面是研究项目的一般步骤，可以了解推论如何进行。我们仍然以青少年对母亲工作的态度作为案例。

事件可能发生的次序如下：

1.研究者从母亲工作的青少年和母亲不工作的青少年中选择代表性样本。样本选择的标准是样本能够代表总体。

2.每一个青少年要进行一次测试来评价他或她的态度。接着计算群体的均值并使用一些检验方法来比较。

3.可以得出的结论是成绩之间的差异是由于偶然性因素引起（也就是母亲工作之外的一些因素是差异的原因），或是由于群体之间“真实的”以及统计显著的差异引起（也就是由于母亲工作）。

4.可以得出的结论是从中选择样本的总体中母亲就业和青少年态度之间的关系。换句话说，基于样本数据分析所得结论进行的推论是关于青少年总体的。

如何选择检验方法

上面的第三步让我们提出这个问题：我如何选择恰当的统计检验来确定群体之间的差异是否存在？检验方法太多了，你必须决定使用哪一种以及如何使用。如何使用哪一

种检验的最好的学习方式是成为有经验的统计学者——选修过这个领域的多门课程而且参与了大量研究。经验永远是最好的老师。实际上没什么真正的学习选用哪一种以及何时使用的方式,除非你具有真实的实际应用这些工具的机会。因此上这门课你就在学习如何使用这些特定的工具。

因此,为了实现我们的目的并开始学习,我们建立了各种工具的简易流程图(或速查图),如图 9.1 所示。你必须知道你在做什么,这样选择正确的统计检验就不完全是随意的决定,当然也是很好的开始学习的起点。

不要认为图 9.1 可以替代你学习不同的检验应该何时使用的需要。这里的流程图只是帮助你开始学习。



这真的很重要。我刚才已经提到选择合适的统计检验方式不是很容易。最好的学习如何做的方式就是实际地去做,这意味着实践和学习更多的统计课程。我们提供的简单的流程图能够有帮助,但是使用过程中要谨慎。当你进行选择时,最好向你的教授或者其他对内容很了解并感觉比你更自信(他知道更多)的人咨询。

如何使用流程图



1—假定你是刚入门的统计人员(实际上也是),对显著性检验有一定的了解,但是对于何时使用哪一种很迷惑。

2—回答流程图上部的问题。

3—依据对流程图每一个问题的回答进行选择,直到流程图的末端。那就是你应该选用的统计检验。这没有火箭科学那么难,而且经过一些实践(你可以通过本书的这一部分来实践),你就能够快速有效地选择恰当的检验。本书这一部分的每一章都会有类似图 9.1 中所见的流程图,让你经过特定的步骤选择应该使用的统计检验。



图 9.1 中的简易流程图是否包含所有的统计检验？当然不是全部。统计检验方法大约有上百种,图 9.1 只列出了最常用的。而且你熟悉了你所在领域的研究之后,就会固定于其中几种检验方法。

显著性检验介绍

推论统计的优势就是可以依据样本的信息得出关于总体的结论。进行推论的最有用的工具之一就是统计显著性检验,显著性检验可以依据所提问题的性质和零假设的形式而应用于不同的情形。

例如,你是否想了解两个群体之间的差异,如男孩的某些考试成绩是否与女孩的成绩有显著差别？或者是两个变量之间的关系,如一个家庭中孩子的数量和智力测试的平均成绩？这两个案例需要不同的方法,但是两者最后都会使用特定的统计显著性检验对零假设进行检验。

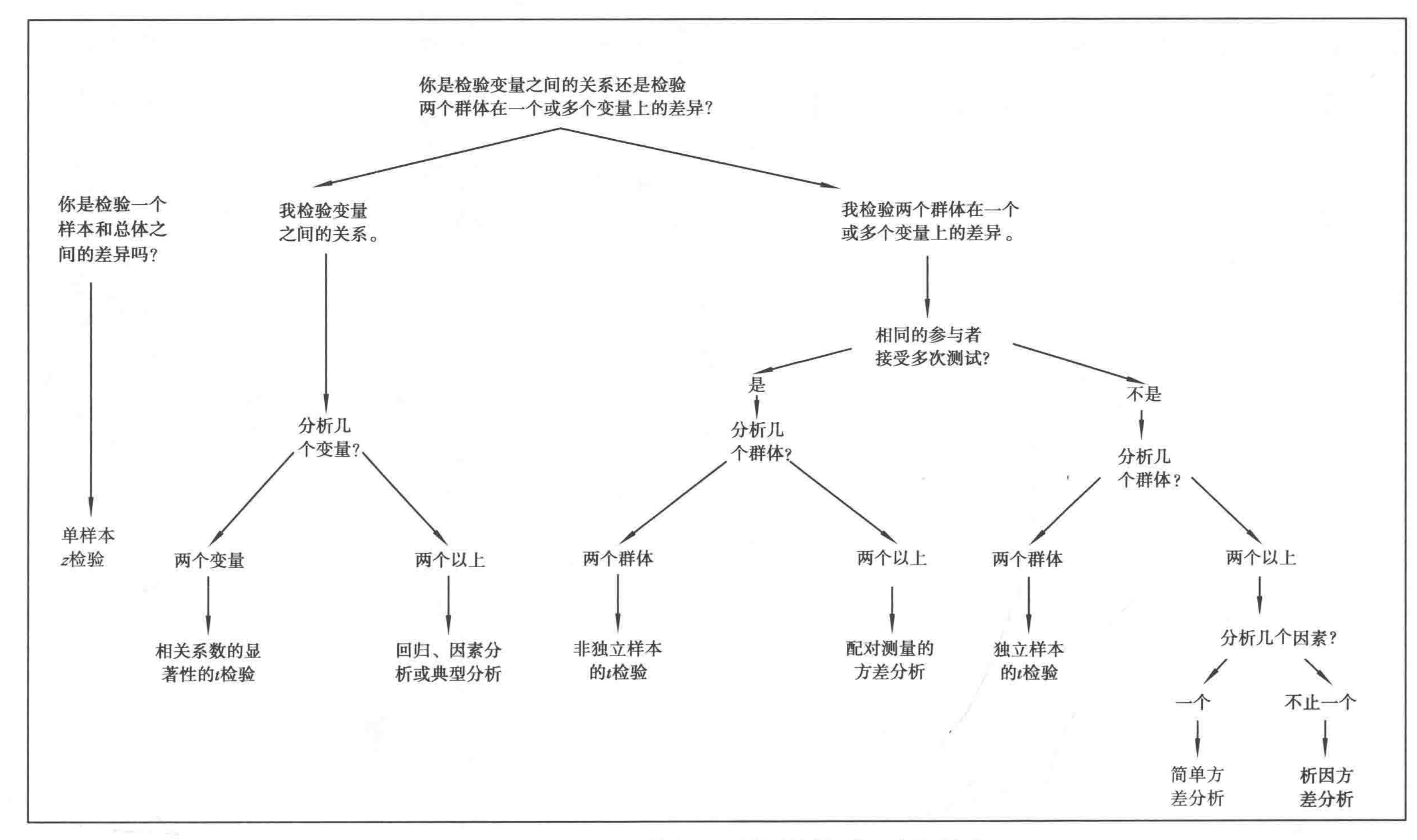


图9.1 选择统计检验方法的快速流程（并不总是最好的）

如何进行显著性检验:计划

显著性检验是基于这样的事实,每一类型的零假设都与特定的统计类型联系在一起。而每一种统计类型与特定的分布联系在一起,你要比较从样本获得的数据的分布。依据样本特征与检验的分布特征的比较你可以推断样本特征是否不同于预期的随机分布特征。

下面是任何零假设进行统计检验时需要采用的一般步骤。这些步骤也是第Ⅳ部分各章的模式。

1.零假设的陈述。你是否还记得零假设的陈述形式是等式?零假设是假定没有其他信息可用于做出判断的情况下事件的“真实”状态。

2.设置零假设的风险水平(或者显著水平,或第一类错误)。任何研究假设都要设置你可能犯错的特定风险水平。第一类错误越小(如0.01与0.05相比),你承担的风险越小。没有假设检验是完全没有风险的,因为你永远不会知道两个变量之间的“真实”关系。要记住按惯例第一类错误设置的概率水平是0.01或0.05。Excel和其他软件可以给出准确的概率水平。

3.选择恰当的检验统计量。每一个零假设伴随着特定的检验统计量。在本书的这部分你可以知道什么样的检验与什么样的问题类型相联系。

4.计算检验统计量。检验统计量(也叫实际值)是特定的统计检验的结果。例如,两个群体的平均值之间差异的显著性,相关系数与0值之间差异的显著性,两个比例值之间差异的显著性,都需要进行统计检验,得到一个具体的数值。

5.使用特定统计量的统计临界值表示确定拒绝零假设需要的值。每一个检验统计量(同时考虑群体规模和愿意承担的风险)都有相应的**临界值**。这个值是零假设,是真实的情况下你预期的拒绝零假设的统计检验值。你应当知道,现在有越来越多的计算机软件可以得到统计检验值对应的准确概率。有了这样的替代方式,这一步及下一步的计算就不那么必要了。

6.比较实际值和临界值。这是关键的一步。就是比较通过检验统计量获得的值(你计算所得的值)与你预期的随机情况下对应的值(临界值)。

7.如果实际值大于临界值,不能接受零假设。也就是说,零假设的等式陈述(反映偶然性)不是我们发现的差异的最有力解释。这也是推论方法表现出其优点的地方。只有实际值大于随机状态对应的值(也就是统计检验的结果不是随机变动的结果)你才可以说你发现的差异不是随机结果,且零假设的等式陈述不是解释你发现的差异的最有力解释。相反,差异一定是由于处理因素引起。

8.如果实际值没有超过临界值,零假设是最有力的解释。换句话说,如果你不能表明你发现的差异是由于偶然因素之外的因素(如处理因素)引起,那么差异一定是由于偶然因素或者其他你没有控制的因素引起的。此时零假设是最好的解释。

一张图胜过千言万语

在图9.2中你可以看到我们刚才学习的八个步骤的表述。这是实际值和临界值比较时不同情况的图形表示。在这个案例中,显著水平设置为0.05或说5%。也可以设置为0.01或1%。

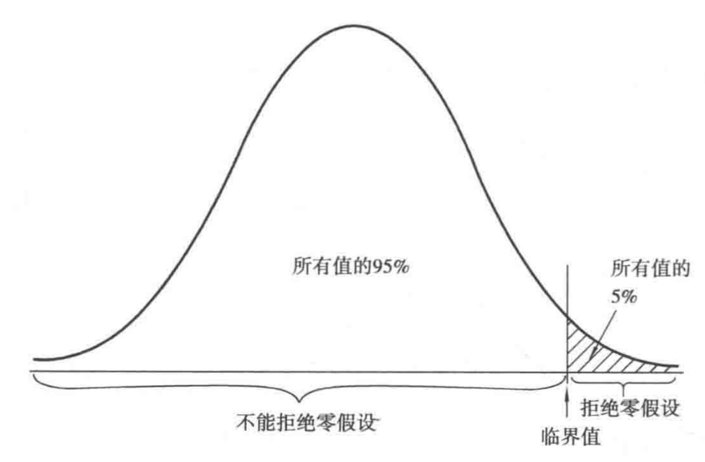


图 9.2　比较实际值和临界值，并作出拒绝或者接受零假设的决定

1.整个曲线表示基于特定零假设的所有可能的结果，如两个群体之间的差异或相关系数的显著性。

2.临界值是这样的点，即超过这个点的实际结果如此稀少，因此可以推断实际结果不是由于偶然因素而是由于其他因素引起的。在这个案例中，我们定义的稀少程度是发生的可能性不足 5%。

3.如果结果表示实际值落在临界值的左边（就是小于极值），结论就是零假设是观察到的差异的最有力的解释。换句话说，实际值落在这个范围（曲线下面积的 95%），我们只能推测结果是由于偶然因素引起的。

4.如果实际值落在临界值的右边（就是大于极值），结论就是研究假设是观察到的差异的最有力的解释。换句话说，实际值落在这个范围（曲线下面积的 5%），我们只能推测结果是由于偶然因素之外的因素引起的。

置信区间——更确信

你现在已经了解不同的结果对应着不同的概率水平——这是本章一直讨论的内容。现在我们使用不同的方式来讨论相同的问题，并将这个新的方式叫作置信区间。

置信区间在样本统计值的基础上对总体参数值进行范围的估计。例如，我们知道 20 个三年级学生（从整个学区的三年级学生中抽取）的拼写成绩的均值，我们多大程度上确信总体的均值将会落在两个数值之间？例如，一个 95%的置信区间是正确的（定义为样本统计值可以代表总体参数）。

你已经知道一个原始数值落在±1.96 的 z 值或者标准差之间的概率是 95%，对吧？（回顾第 8 章相关内容）或者原始数值落在±2.56 的 z 值或者标准差之间的概率是 99%。如果我们使用对应 z 值的正负原始数据，我们就得到一个置信区间。

让我使用一些真实的数值来解释。

例如，一个 100 人构成的六年级学生随机样本的拼写成绩均值是 64（一共 75 个单词）标准差是 5。我们可以在多大程度上推测整个六年级学生总体的拼写成绩均值？

95%的置信区间等于：

$$64\pm 1.96(5)$$

换句话说，数值范围是 54.2 到 73.8，因此你至少可以 95%地确信整个六年级学生的平均拼写成绩落在这两个数值之间。

如果你想更加确信，就可以计算 99%地置信区间：

$$64 \pm 2.56(5)$$

数值范围是 51.2 到 76.8，因此你可以 99%的确信总体均值落在这两个数值之间。

为什么随着你确信的概率水平（从 95%到 99%）提高置信区间扩展？因为置信区间越大（在这个案例中，95%的置信水平对应的区间范围是 73.8−54.2＝19.5，99%的置信水平对应的区间范围是 76.8−51.2＝25.6），包括可能的结果的范围就越大，你就可以更加确信。这是不是很酷？

小　结

现在你已经明确地了解了显著性概念如何应用，余下的工作就是将显著性概念应用到不同的研究问题中。这也是下一章开始的内容，也会在本书这一部分的大部分章节中继续。

练　习

1.为什么显著性是研究和应用推论统计的重要概念？

2.下面的陈述有什么问题？

a.第一类错误是 0.05，意味着 100 次中有 5 次我会拒绝真实的零假设。

b.将第一类错误设置为 0 是可能的。

c.第一类错误的概率水平越小，结果越好。

3.检验研究假设的显著性时为什么考虑随机因素？

4.给定以下信息，你会接受还是拒绝零假设？（显著水平定义为 0.05）。解释你的理由。

a.零假设是个人所听的音乐的类型与犯罪率之间没有关系，$p<0.05$。

b.零假设是学生咖啡消耗量和学习成绩之间没有关系，$p=0.62$

c.研究假设是个人工作时间和工作满意度之间负相关，$p=0.51$

5.为什么研究假设在 0.01 显著水平下检验比在 0.05 显著水平下更难找到显著性结果？

6.为什么我们认为是“没有拒绝”零假设而不是接受零假设？

7.临界值表示什么？

8.实际值表示什么？

9.如果你想分析两个独立群体之间是否存在差异，使用图 9.1 如何选择？

10.在图 9.2 中，在曲线的右侧有一个阴影区域。

a.这个区域表示什么？

b.如果你在更严格的水平（例如是 0.01 而不是 0.05）检验研究假设，这个区域是变大还是变小，为什么？

10 只有一个——单样本 Z 检验

本章你会学到什么

- 何时使用单样本 Z 检验
- 如何计算 z 值
- 如何解释 z 值并理解 z 值的意义

单样本 Z 检验介绍

睡眠缺乏会导致各种问题，如焦躁、疲劳，甚至死亡。因此你可以想象医疗保健专家如何重视让他们的病人得到足够的睡眠。如果你已经生病并真的需要睡眠带来治疗和恢复的效果，就更是如此。约瑟夫·卡佩莱里（Joseph Cappelleri）博士和他的同事研究患有纤维肌痛的病人的睡眠困难，评估用作测量睡眠问题的工具——医疗效果研究（MOS）睡眠量表是否有用。同时也完成了其他的分析，包括实验组和控制组的比较，重要的分析（对我们来说）是比较参与者的 MOS 得分和全国 MOS 标准。样本值（在这项研究中参与者的 MOS 值）和总体值（标准）之间的比较，需要使用单样本 Z 检验。研究者有什么发现？MOS 睡眠量表得分统计上很低（$p<0.001$），意味着研究参与者睡眠不是很好。这意味着被检验的样本与一般总体之间没有相同的特征（至少在睡眠测量方面）。换句话说，"样本平均值和总体平均值之间相等"这个零假设是不能接受的。

为什么使用单样本 Z 检验。卡佩莱里和他的同事想知道样本值是否和使用相同测量获得的总体值有区别。实际上他们比较了样本值和总体参数值，并看他们是否能得出样本可以或者不可以代表总体的结论。

想了解更多？查阅卡佩莱里、布什马金等人（Cappelleri, J. C., Bushmakin, A. G., McDermott, A. M., Sadosky, A. B., Petrie, C. D., & Martin, S.）2009 年发表在《健康和生活质量》（*Health and Quality of Life Outcomes*）第 7(54) 期第 766—770 页上的文章"评估纤维肌痛病人睡眠质量的单一项目量表的心理测量特征"（Psychometric properties of a single-item scale to assess sleep quality among individuals with fibromyalgia）。

智慧和知识之路

如何使用图 10.1 也就是第 9 章介绍的流程图来选择合适的检验统计量，即单样本 z 检验？沿着图 10.1 指示的步骤即可。这个过程很容易（但并不总是这么容易），因为这是

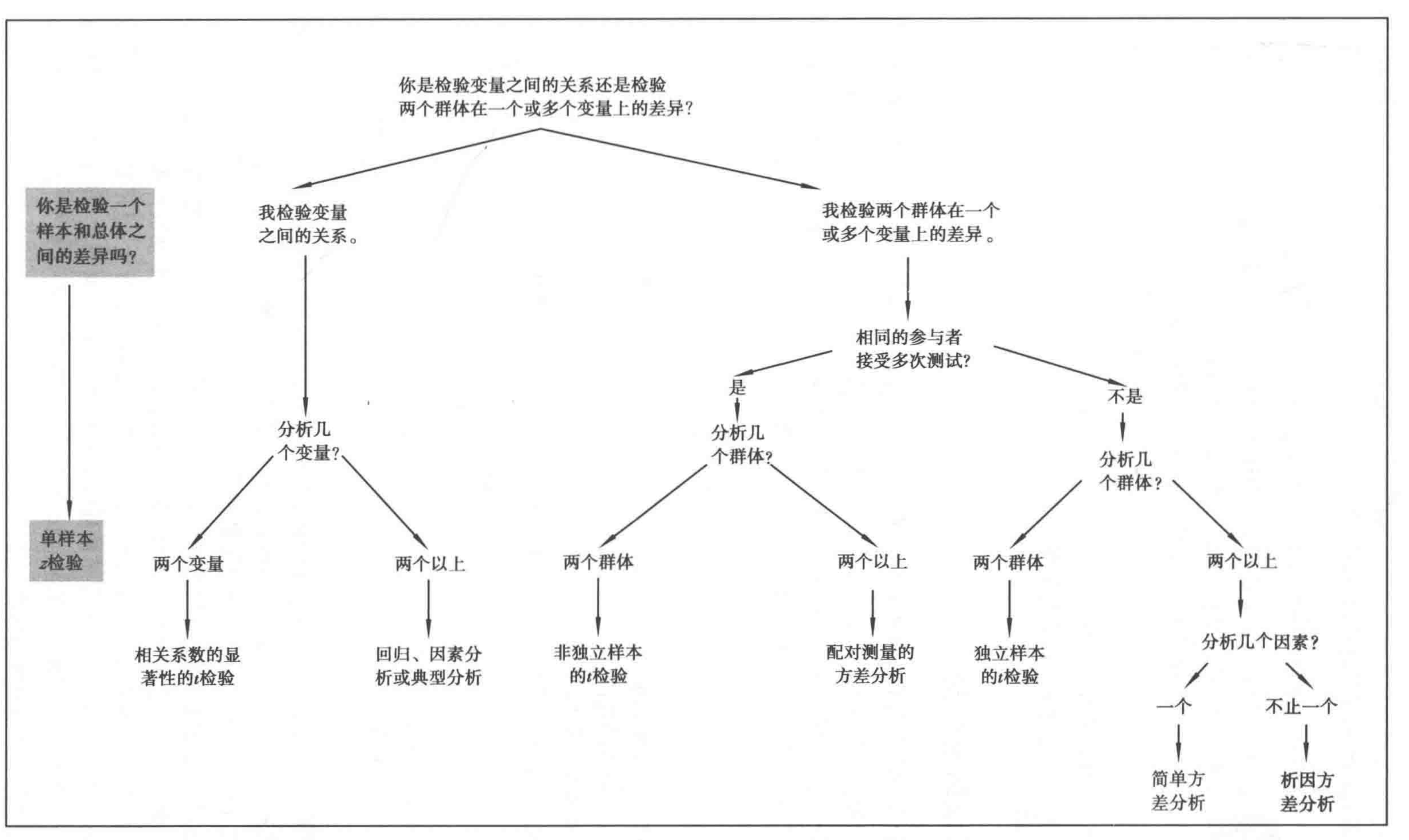
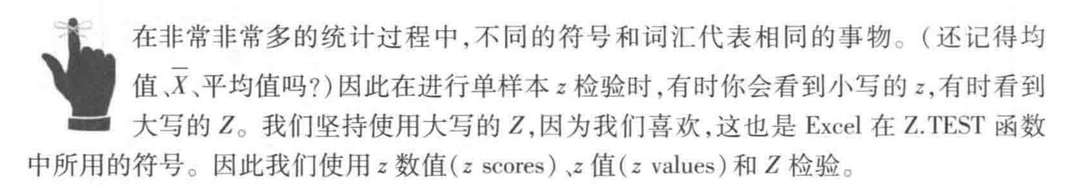


图10.1 确定单样本z检验是合适的统计检验方法

本书第四部分唯一只有一个群体的推论统计。接下来学习的许多内容需要你回到第 8 章和标准分部分，而你对这些内容已经有所掌握。



在非常非常多的统计过程中，不同的符号和词汇代表相同的事物。（还记得均值、$\overline{X}$、平均值吗？）因此在进行单样本 z 检验时，有时你会看到小写的 z，有时看到大写的 Z。我们坚持使用大写的 Z，因为我们喜欢，这也是 Excel 在 Z.TEST 函数中所用的符号。因此我们使用 z 数值（z scores）、z 值（z values）和 Z 检验。



1.我们检验一个样本和总体之间的差异。
2.合适的统计检验是单样本 Z 检验。

计算检验统计量

公式 10.1 是计算单样本 Z 检验值的公式。要记住我们是检验一个样本的均值是否属于或者代表一个总体的均值。样本均值（$\overline{X}$）和总体均值（μ）之间的差异构成了 z 值的分子。分母是一个误差项，就是均值标准误差（SEM）。使用均值标准误差（这儿的关键在于"标准"）可以在此允许我们（如第 9 章所示）使用 z 值表来确定结果出现的概率。

$$Z = \frac{\overline{X} - \mu}{\text{SEM}} \tag{10.1}$$

其中：$\overline{X}$ 是样本的均值，
μ 是总体的均值，
SEM 是均值标准误差。

使用下面的公式来计算公式 10.1 中的均值标准误差 SEM。

$$\text{SEM} = \frac{\sigma}{\sqrt{n}} \tag{10.2}$$

其中：σ 是总体的标准偏差，
n 是样本规模。

Tech Talk

均值标准误差是从总体中抽取的所有可能的样本的均值的偏差。样本均值的数值要比单个样本的均值更接近总体均值。样本均值是由几个样本平均所得，部分消除了样本的不均匀性并降低了偶然误差。因为不可能计算所有可能的均值，因此均值标准误差是我们可以进行的最好估计。如果我们的抽样是完美的，样本均值和总体均值之间的差异将会是 0，是吧？是的。另一方面，如果从总体中抽取样本的过程没有正确（随机地，代表总体地）进行，那么所有样本的标准差将会非常大，是不是？是的。所以我们的确在尽力做，但是我们似乎不能抽取完美的样本——总是存在一些误差，均值标准误差反映的是对所有的样本均值的总体来说，这个误差值大概是多少，这就是均值标准误差。当然还有其他量数的标准误差。

举例说明

麦克唐纳博士认为他的地球科学学生团队是特别的(在好的方面),他想知道这个学生团队的平均成绩是否落在过去 20 年上过地球科学课程的这个更大的学生群体的平均成绩范围内。因为麦克唐纳博士做了详细的成绩记录,他知道他现在的 36 人学生团队以及过去和现在总共注册的 1 000 人的更大的学生团队的均值和标准差。下面是他的数据。我们用于计算 z 值的数据下面有一横。

	规模	均值	标准差
样本	<u>36</u>	<u>100</u>	5
总体	1000	<u>99</u>	<u>2.5</u>

下面是计算 Z 检验统计量的重要的八个步骤。

1.陈述零假设和研究假设。零假设是样本平均值等于总体平均值。如果零假设没有被拒绝,那么样本是总体的代表。如果零假设被拒绝而接受研究假设,那么样本平均值不同于总体平均值。

零假设是:

$$H_0: \bar{X} = \mu \tag{10.3}$$

研究假设是:

$$H_1: \bar{X} \neq \mu \tag{10.4}$$

2.选择零假设的风险水平(或者显著水平或者第一类错误)。风险水平或者显著水平或者第一类错误(是否还有其他名称?)是 0.05。这个数字完全由研究者决定。

3.选择合适的检验统计量。使用图 10.1 所示的流程图,我们决定使用的检验是单样本 Z 检验。

4.计算检验统计量(也叫作实际值)。现在是时候代入数值进行一些计算。公式 10.1 是 z 值计算公式。将具体的数值代入,首先是在公式 10.5 计算 SEM,公式 10.6 计算 z 值。所有的数值都来自上表。代入数值之后,我们得到下面的结果。

$$\text{SEM} = \frac{2.5}{\sqrt{36}} = 0.42 \tag{10.5}$$

$$z = \frac{100 - 99}{0.42} = 2.38 \tag{10.6}$$

依据上表中的数据,z 值用于比较样本均值和总体均值,为 2.38。

5.使用特定统计量的临界值表来决定拒绝零假设需要的数值。我们需要查询附录 B 的表 B.1,表中列出了与特定 z 值对应的概率。查阅这个案例中拒绝零假设需要的临界值。我们可以使用表 B.1 中的数值,通过我们随机预期的值(表中的数值或者临界值)和我们实际观察的数值(实际值)比较,来发现两个均值是否相互“属于”。

在第 9 章几个案例中我们就是这么做的。从第 9 章的学习内容来看,我们知道 z 值 +1.96相对应的概率是 0.025。如果我们考虑到样本均值可能比总体均值更大或者更小,我们需要考虑分布的两侧(在±1.96 范围内),那么总的第一类错误是 0.05。

6.比较实际值和临界值。实际值是 2.38，在 0.05 的置信水平下，样本规模是 36，对零假设的检验来说，临界值是±1.96。这个值是转折点，这个点以内随机因素是解释样本均值和总体均值不同的最有吸引力因素，如果在两个方向上超过这个值（记住我们的研究假设是没有方向的，这是一个双侧检验）意味着我们需要提供一个解释，解释为什么样本均值不同于总体均值。

7.和 8.做出决定。

如果实际值远远大于临界值（记住图 9.2），零假设不能被接受。如果实际值没有超过临界值，零假设就是最有吸引力的解释。在这个案例中，实际值（2.38）超过了临界值（1.96），实际值足够大，我们可以说麦克唐纳博士班上的 36 个学生的平均成绩不同于之前也上过这门课的 1 000 个学生的平均成绩。如果实际值小于 1.96，我们得出的结论就是样本和总体的考试成绩没有差异，样本的 36 个学生获得的成绩和之前的 1 000 个学生总体基本上是一样的。

如何解释 $z=2.38, p<0.05$？

- z 表示我们使用的检验统计量。
- 2.38 表示实际值，使用这一章之前给出的公式计算。
- $p<0.05$（这个不等式非常重要）表示零假设的任何检验的概率小于 5%，也就是样本均值不同于总体的均值。

使用 Excel 的 Z.TEST 函数计算 z 值

这是你第一次使用 Excel 进行 Z 检验，来看一个特定的值是否属于数值总体。作为 Excel 工具包中的工具，这个工具需要一些很小的改变，需要使用不同的数据信息。所用的数据与上面的案例给出的一样，只是 Excel 要求总体。因此我们录入 100 个数值，总体的均值是 X，数据见第 10 章数据集 1。均值是 74.59，我们检验的数值是 76。换句话说就是数值 76 是否代表其总体？

1.单击［公式］→［其他函数］→［统计］→［Z.TEST］。这样就可以看到 Z.TEST 函数对话框，如图10.2所示。就我们的目的来说，我们将函数放在单元格 C1。

图 10.2　Z.TEST 函数对话框

2.选择包含数据的列(从单元格 A1 到单元格 A101)。

3.输入将要检验的数值(X= 76)来确定这个数值是否属于总体。你可以看到图 10.3 所示的已完成的 Z.TEST 对话框。

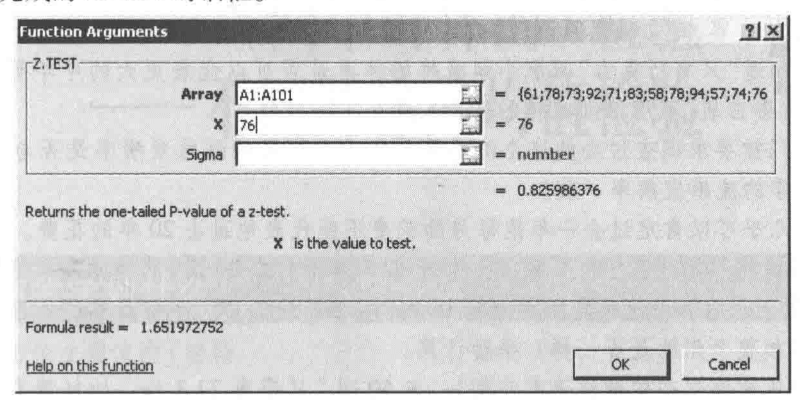


图 10.3 已完成的 Z.TEST 对话框

4.单击确定,如图 10.4 所示,返回的数值是 0.826。这意味着特定的值(76)属于数值总体(第 10 章数据集 1)的可能性非常高。或者换句话说就是特定的数值(76)与从中选取的数值(总体)之间没有显著性的差别。用案例结果来说,就是这个学生的成绩不是很“特别”。

C1 =Z.TEST(A1:A101,76)

	A	B	C	D	E	F
1	61		0.826			
2	78					

图 10.4 数值 76 和第 10 章数据集 1 的 Z.TEST 结果

更多 Excel

有一点非常重要,需要记住:Z.TEST 函数不提供进行检验或者评价显著性的数值,我们在本章之前的手动计算时提供了这个数值。实际上 Z.TEST 函数提供了这个数值发生的可能性。因此,与本章之前使用的手动计算方法不同,Z.TEST 函数返回的是这个数值作为总体中的一个数值的概率,而不是与这个原始数值相对应的 z 值。

小 结

单样本 Z 检验是重要的单样本推论统计,这也是为什么首先介绍这个检验方法来解释什么是检验以及如何应用。好消息是我们在学习更复杂的分析工具时采取的大多数步骤与我们现在学习的步骤一样。下面我们会学习非常常用的推论检验(也是我们这里学习的 Z 检验的扩展)——两个群体之间均值的简单 t 检验。

练　习

1.什么时候适合使用单样本 Z 检验？

2.为什么 Z 检验中使用大写字母 Z，在什么情况下与 z 或者标准分相同？

3.按照下面的内容，建立研究假设。

a.鲍勃想知道"只有巧克力"减肥小组减掉的体重是否可以代表更大的中年男性总体参与"只有蛋白质"减肥活动减掉的体重。

b.保健部门被要求调查过去的这个流感季节每千个市民的流感发病率是否与过去的 50 个流感季的流感发病率一致。

c.布莱尔几乎可以肯定过去一年他每月的花费不能代表他过去 20 年的花费。

4.堪萨斯巴特菲尔德地区学校系统内上个流感季（4 个月或 20 周）的流感发病案例是每周 15 例。整个州的每周流感发病案例是 16 例，标准差是 2.35。巴特菲尔德地区孩子的流感发病率和整个州的是否一样？手动计算。

5.布兰达在伍德维尔高中游泳速度非常快，游 50 码* 只需要 22.3 秒。她打算离开伍德维尔所在的县去一个新的社团，这个社团的州记录是 21.5 秒。使用第 10 章数据集 2，这个数据集就是州竞赛中全州范围内 1 000 个游泳选手的 50 码成绩。使用 Excel 的Z.TEST 函数来分析布兰达在她的新游泳片区中的位置。

6.多大的差异才是足够大的差异？薯片天才正在研发一种没有土豆的薯片，而且保质期很长。新的薯片没有脂肪，味道也不错。为了检验新的薯片是否和真的薯片的味道一样好，品尝专家让消费者使用一个 1 到 10 的量表来给薯片打分。第 10 章数据集 3 是 500 个品尝者对真的薯片的打分。消费者给新薯片所打分数的平均值是 5.1。是否值得投资 1 亿来建立新薯片的市场（即，新薯片的得分是否与真薯片一样）？

---

* 1 码=0.914 4 米

两个群体的 t 检验 ——不同群体的均值检验

11

本章你会学到什么

- 何时使用独立均值的 t 检验
- 如何计算观察值 t 值
- 如何使用 T.TEST 函数
- 如何使用分析工具库的 t 检验工具计算 t 值
- 解释 t 值，理解 t 值的含义

独立样本 t 检验介绍

即使饮食紊乱因为其严重性得到关注，也很少进行症状的盛行程度和严重性的跨文化比较研究。约翰·舍斯泰特、约翰·舒马克与纳特沃特(John P. Sjostedt, John F. Schumaker, & S.S. Nathawat)对 297 个澳大利亚大学生和 249 个印度大学生进行了这项比较研究。每个学生都参加了饮食态度测试和 Goldfarb 肥胖恐惧量表测试，然后比较群体得分。就澳大利亚学生和印度学生的均值比较而言，印度学生的两个测试的得分都高于澳大利亚学生。饮食态度测试的结果是 $t_{(544)}=-4.19, p<0.000\ 1$；Goldfarb 肥胖恐惧量的测试结果是 $t_{(544)}=-7.64$, $p<0.000\ 1$。

那么，结果意味着什么？接着往下看。

为什么 t 检验用于独立均值？舍斯泰特和他的同事的兴趣在于发现两个相互独立的群体在一个(或多个)变量的平均值上是否有差异。我们所指的独立性的含义是两个群体在任何方面都不相关。研究中的每一个参与者只接受一次测试。研究者采用独立均值的 t 检验，依据对每一个结果变量得出的结论，两个群体之间的差异在等于和小于 0.000 1 的显著水平下是显著的。这么小的第一类错误意味着两个群体得分的差异是由群体成员本身之外的偶然因素引起的几率非常小，在这个案例中群体本身的因素是指民族、文化或者种族。

是否想了解更多？查阅约翰·舍斯泰特、约翰·舒马克与纳特沃特(Sjostedt, J.P., Schumaker, J.F., & Nathawat, S.S.)1998 年发表在《社会心理学杂志》(*Journal of Social Psychology*)138 期第 3 卷 351-357 页上的文章“印度大学生和澳大利亚大学生的饮食紊乱调查(Eating disorders among Indian and Australian university students.)”。

智慧和知识之路

下面介绍如何使用图 11.1 所示的第 9 章已经介绍的流程图选择合适的统计检验，也就是独立均值的 t 检验。沿着图 11.1 中所示的步骤完成即可。



1—研究澳大利亚学生和印度学生之间的差异。
2—每一个参与者只被测试一次。
3—有两个群体。
4—合适的统计检验是独立均值的 t 检验。



几乎每一个统计检验都有特定的假设支持检验的使用。例如，t 检验的一个主要假设是两个群体中每个群体的变异性的量是相等的。这是方差齐性假定。如果样本规模足够大就会破坏这个假定，小样本或者假定的破坏都可能导致自相矛盾的结果和结论。你不需要太担心这些假定，因为这些内容已经超过了本书的范围。不过，你应该知道，虽然这样的假定很少被破坏，但这种可能确实存在。

计算检验统计量

公式 11.1 是计算独立均值 t 检验中 t 值的公式。公式的分子是均值之间的差。群体内和群体之间的变化的数量构成分母。

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right]\left[\frac{n_1 + n_2}{n_1 n_2}\right]}} \tag{11.1}$$

其中：$\overline{X}_1$ 表示群体 1 的均值，

$\overline{X}_2$ 表示群体 2 的均值，

n_1 表示群体 1 中参与者的数量，

n_2 表示群体 2 中参与者的数量，

s_1^2 表示群体 1 的方差，

s_2^2 表示群体 2 的方差，

公式中没有新的内容，重要的只是代入正确的值。

一个例子

下面的数据是为帮助老年痴呆症患者记住日常生活秩序而设计的项目中患者记住的单词的数量。群体 1 使用视觉教学，群体 2 使用视觉教学和发声训练。我们使用这些数据计算下面案例的检验统计量。

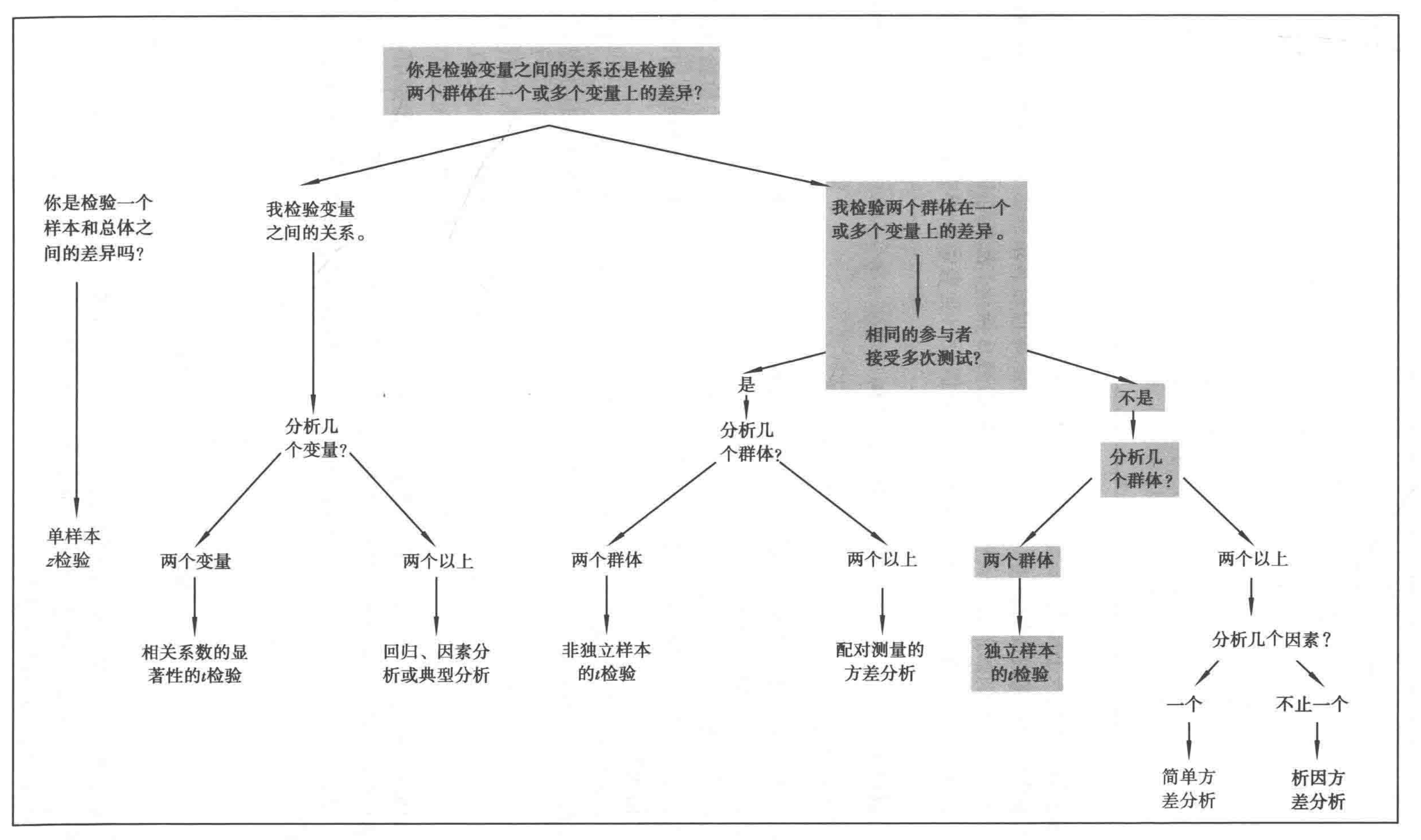


图11.1 确定 t 检验是合适的统计检验方法

群体 1			群体 2		
7	5	5	5	3	4
3	4	7	4	2	3
3	6	1	4	5	2
2	10	9	5	4	7
3	10	2	5	4	6
8	5	5	7	6	2
8	1	2	8	7	8
5	1	12	8	7	9
8	4	15	9	5	7
5	3	4	8	6	6

下面是著名的八个步骤和 t 检验统计量的计算。

1.零假设和研究假设的表述。如公式 11.2 所示，零假设表示群体 1 和群体 2 的均值之间没有差异。就我们的目的而言，研究假设(公式 11.3 所示)表示两个群体的均值之间有差异。研究假设是双侧的无方向假设，因为研究假设只是表示差异存在，而没有特定的方向。

零假设是：
$$H_0: \mu_1 = \mu_2 \tag{11.2}$$

研究假设是：
$$H_1: \overline{X}_1 \neq \overline{X}_2 \tag{11.3}$$

2.设置零假设的风险水平(或显著性水平，或第一类错误)。风险水平或第一类错误或显著性水平(或其他的称谓)是 0.05，这完全由研究者决定。

3.选择合适的检验统计量。使用图 11.1 所示的流程图，我们确定合适的检验方法是独立均值的 t 检验。因为这两个群体相互独立，所以不是非独立均值的 t 检验(初学者常犯这个错误)。

4.计算检验统计值(也叫作实际值)。现在我们代入观察值并进行计算。公式 11.1 是 t 值公式。代入具体的值之后，我们就得到等式 11.4(我们已经计算了均值和标准差)：

$$t = \frac{5.43 - 5.53}{\sqrt{\frac{(30-1)\times 3.42^2 + (30-1)\times 2.06^2}{30+30-2} \times \frac{30+30}{30\times 30}}} \tag{11.4}$$

代入具体的数值之后，公式 11.5 表示我们如何得到最后的值-0.137 1。因为是用一个较小的值(群体 1 的均值是 5.43)减去一个较大的值(群体 2 的均值是 5.53)，所以这个值是负值。要记住一点，检验是无方向的，而且任何差异都被假设了，所以差异的符号也是无意义的。

$$t = \frac{-0.1}{\sqrt{\frac{339.20 + 123.06}{58} \times \frac{60}{900}}} = -0.14 \tag{11.5}$$



如果讨论的是无方向的检验，你会发现 t 值是用绝对值来表示的，类似于 $|t|$，忽略了所得值的符号。你的老师甚至会说，t 值的符号在有方向的检验中才有意义，无方向的检验就不必在意了。

5.使用特定的统计量的临界值分布表确定拒绝零假设需要的值。现在我们需要查阅附录 B 的表 B2，表 B2 列出了 t 检验的临界值。我们可以使用这个数据分布表来看两个独立均值是否不同，这要通过比较预期的随机数值（分布表中的值或者临界值）与观察到的值（实际值）来实现。

我们的第一个任务是确定**自由度**（degrees of freedom，df），自由度近似于样本规模。对目前选定的检验统计量来说自由度是$(n_1-1)+(n_2-1)$。因此对每一个群体来说，就是将两个样本的规模加起来然后减去 2。在这个案例中就是 30+30−2＝58。这是就这类检验统计量的自由度，但对其他统计量不合适。



不论你使用什么统计检验，自由度的意义几乎都是相同的。但是计算特定检验的自由度的方式每个老师或者每本书可能会不同。我们告诉你上面检验正确的自由度是$(n_1-1)+(n_2-1)$。但是一些教师认为应该使用两个样本规模中更小的那个（这是你可以考虑的更保守的方式）。

使用这个数字（58）、承担你愿意承担的风险水平（早先定义的0.05）以及进行双侧检验（因为研究假设没有方向），你就可以使用 t 检验表来查找临界值。对于显著水平为 0.05、自由度为 58 的双侧检验来说，拒绝零假设需要的值就是……哦！在分布表中没有自由度 58 这个数值！你该做什么？如果选择对应自由度 55 的值，你会显得保守，因为你使用了小于现有样本规模的样本对应的值（临界值 t 值会变大）。

如果你选择对应自由度 60（最接近 58 的值）的值，你会更接近总体规模，但是选择 60 相对于 58 而言更为随意一些。虽然统计学家对于这种情况该怎么做的观点不同，我们通常选择最接近于实际样本规模的值。因此在显著水平为 0.05、自由度为 58 的情况下拒绝零假设需要的值是 2.001。

6.比较实际值和临界值。实际值是−0.14，拒绝零假设（也就是群体 1 和群体 2 取得的成绩没有差异）的临界值是 2.001。临界值 2.001 表示对应于这个值，在愿意承担的风险水平为 0.05、每个群体 30 个参与者的情况下，随机因素是对两个群体之间观察到的任何差异的最有力的解释。

7.和 8.做出决定。

现在我们该做出决定了。如果实际值大于临界值（参照图 9.2）就不能接受零假设。如果实际值没有超过临界值，零假设就是最有力的解释。在这个案例中，实际值（−0.14）没有超过临界值（2.001）——这个值不够大，我们不能说群体 1 和群体 2 之间的差异是由于随机因素之外的因素引起的。如果实际值等于或者大于2.001，就如投硬币试验中 10 次有 8、9 或 10 次都得到正面一样，这个值大到我们不能相信除了随机因素之外其他因素没有发挥作用。在投硬币试验中，是不均匀的硬币；在这个案例中，一定有其他更好的方法来向老年人教授记忆能力。

那么是什么引起两个群体之间这么小的差异？如果我们继续现在的讨论，那么可以说差异是由于抽样误差或者对参与者成绩的微小变化的四舍五入误差引起的。最重要的是我们可以确信（当然不是 100%确信）不是由于任何特定的因素使得一个群体或另一个群体取得更好的成绩。

如何解释 $t_{(58)} = -0.14, p>0.05$

- t 表示我们所用的检验统计量。
- 58 是自由度数值。
- −0.14 是实际值，是使用本章之前给出的公式计算所得的值。
- $p>0.05$（p 值是最重要的）说明对零假设的任何检验来说，两个群体的差异来源于随机因素的可能性大于 5%。同时注意 $p>0.05$ 也以 $p=n.s.$的形式表示非显著性。

现在，使用 Excel 的 T.TEST 函数

Excel 没有计算检验两个独立群体之间差异的 t 值的函数。T.TEST 函数计算返回的是数值发生的概率。非常有用，但是如果你需要在报告中使用 t 值，那你就不走运了（实际上你的人生挺走运的，因为分析工具库里有非常不错的相应函数可用）。操作步骤如下。



1.在工作表中按照列录入数据。列的标签是群体 1（group 1）和群体 2（group 2），图 11.2 中你可以看到部分数据。

2.选择你要输入 T.TEST 函数的单元格。在这个例子中，我们希望 T.TEST 函数返回的数值在单元格 D1（选择这个单元格没有什么特别的原因）。

3.现在单击［公式］→［其他函数］→［统计］，在下拉选项中选择 T.TEST 函数，形式如下：

= T.TEST（array1，array2，tails，type），

其中：

array1 是第一个数据集的单元格地址（在这个例子中是 A2：A31），

array2 是第二个数据集的单元格地址（在这个例子中是 B2：B31）。

tails 是 1 还是 2 依据是单侧检验（有方向检验是 1）还是双侧检验（无方向检验是 2）确定。

type 是 1 表示是配对样本 t 检验，2 是两个样本（具有相同方差的独立样本）的检验，3 是方差不相等的两个样本的检验。

	A	B
1	Group 1	Group 2
2	7	5
3	3	4
4	3	4
5	2	5
6	3	5
7	8	7
8	8	8
9	5	8
10	8	9

图 11.2　用于 T.TEST 函数的数据

4.对现在这个例子来说，如图 11.3 所示，完成的 T.TEST 函数表达式如下：

=T.TEST（A2：A31，B2：B31，2，1）

D1　f_x =T.TEST(A2:A31,B2:B31,2,1)

	A	B	C	D	E	F
1	Group 1	Group 2		0.877992		
2	7	5				
3	3	4				

图 11.3　使用 T.TEST 函数计算 t 值的概率

单击确定，你可以看到返回的值是 0.877 992。

需要注意 T.TEST 函数的两个重要的方面。

第一，T.TEST 函数不能计算 t 值。

第二，T.TEST 返回的是由于随机因素（当然假定零假设是真实的）影响而得到的 t 值的似然率（likelihood）。对发生概率是0.88（记住，最大只能达到 1 或者 100%）的 t 值零假设的解释力很高。

> 不管你信不信，在过去，就是本书作者和你的导师还是研究生的时候，只有巨大的大型计算机，而且也没有看到我们今天放在桌子上使用的电脑出现的迹象。换句话说，在统计课上所做的所有计算都是手动计算，这样做的最大好处是，可以帮助你更好地理解这个过程；其次就是没有计算机你也可以做统计分析。因此，如果计算机不能给出你需要的所有数据，就要有些创新。你已经知道获得实际值的基本公式，也有合适的对照表，你会做得很好。

Tech Talk

如果你的第一个数据集（数列 1）和第二数据集（数列 2）的数据点不相同，并且你在"数据类型"下输入 1，那么 T.TEST 函数返回的是#N/A 这样的出错信息。为什么？因为你不能在数据集中得到对等的数据，但是你得到的是匹配（依赖）数据，就如在 12 章所见。

你使用公式计算 t 值和使用函数计算两者之间存在着非常大的差别。要记住的是，如果你手动计算，你必须使用一个数据表来确定临界值，接着与观察值进行比较。如果我们现在使用 T.TEST 函数和本章之后要讨论的分析工具库，这里就没有"$p<$"，这是因为Excel 计算准确的概率。不需要数据表——因为得到的那个数，就是结果发生的概率。

使用奇妙的分析工具库计算 t 值

我们会再一次发现分析工具库可以给我们所有的信息，我们需要这些信息来进行 t 值及其显著性的判断。分析工具库也提供其他的信息，就如你会看到的，非常有用，可以让我们不再进行其他额外的分析。

1.单击［数据］→［数据分析］（Data Analysis），你会看到如图 11.4 所示数据分析对话框。

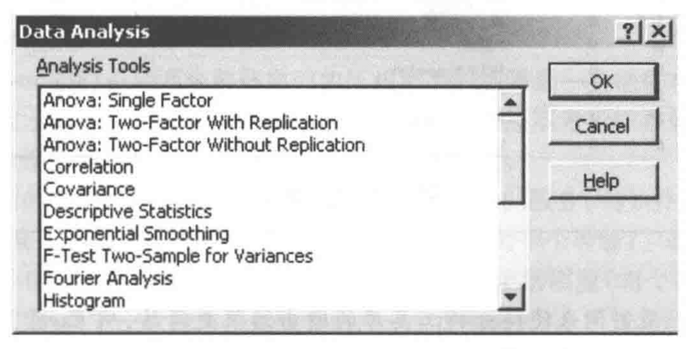


图 11.4　让我们开始使用分析工具库的对话框

2.单击[t 检验：双样本等方差假设](t-Test：Two-Sample Assuming Equal Variances)，然后单击确定，你就会看到如图 11.5 所示的描述统计对话框。

t-Test: Two-Sample Assuming Equal Variances

Input
Variable 1 Range:
Variable 2 Range:
Hypothesized Mean Difference:
Labels
Alpha: 0.05
Output options
Output Range:
New Worksheet Ply:
New Workbook
OK
Cancel
Help

图 11.5　t 检验对话框

3.在[变量 1 的区域](Variable 1 Range)栏，输入第一个数据组的单元格地址。在图 11.2 所示的我们案例工作表中，单元格地址是 A1：A31(包括了标签 Group 1)。

4.在[变量 2 的区域](Variable 2 Range)栏，输入第二个数据组的单元格地址。在图 11.2 所示的我们案例工作表中，单元格地址是 B1：B31(包括了标签 Group 2)。

5.点击[标志](Labels)选项框，这样数据列的标签就会在 Excel 运行的输出结果中出现。

6.点击[输出区域](Output Range)选项，输入你想让结果返回的单元格地址(同一工作表)。在这个案例中，我们将结果放在以单元格 D1 开始的位置。

7.单击[确定](OK)，如图 11.6 所见，你会得到与分析相关的重要的数据的汇总(我们调整了格式，使其看起来很整齐)。在这个图之下是对其中每个统计量是什么的描述清单。

L26

	A	B	C	D	E	F
1	Group 1	Group 2		t-Test: Two-Sample Assuming Equal Variances		
2	7	5				
3	3	4			Group 1	Group 2
4	3	4		Mean	5.43	5.53
5	2	5		Variance	11.70	4.26
6	3	5		Observations	30	30
7	8	7		Pooled Variance	7.98	
8	8	8		Hypothesized Mean Difference	0.00	
9	5	8		df	58	
10	8	9		t Stat	-0.14	
11	5	8		P(T<=t) one-tail	0.45	
12	5	3		t Critical one-tail	1.67	
13	4	2		P(T<=t) two-tail	0.89	
14	6	5		t Critical two-tail	2.00	

图 11.6　数据汇总

更多 Excel

应当花费一点时间来使得分析工具库的输出结果更简洁，从 Excel 复制出来，粘贴(你需要的部分)到另一个文件中。

统计量	描 述
Mean	每个变量的平均数值
Variance	每个变量的方差
Observations	每个数据组观察值的个数
Pooled Variance	两个数据组的方差
Hypothesized Mean Difference	你已经表明的你所预期的差异(返回到对话框)
df	自由度
t Stat	t 统计量的值
P(T<=t) one-tail	单侧检验,随机因素导致的 t 值的概率
t Critical one-tail	单侧检验,你需要超过的临界值(回顾第 8 章临界值内容)
P(T<=t) two-tail	双侧检验,随机因素导致的 t 值的概率
t Critical two-tail	双侧检验,你需要超过的临界值(回顾第 8 章临界值内容)

结 果

分析结果显示虽然群体 2 的确比群体 1 得分高,但是差异不显著。双侧检验 t 值是 -.14,相应的发生概率 p 值是.89。下面是结果的汇总。

	数据组 1	数据组 2
Mean(均值)	5.43	5.53
Variance(方差)	11.70	4.26
Observations(观察数)	30	30
Degrees of freedom(自由度)	58	
t statistic(t 统计量)	-0.14	
p value(p 值)	0.89	

特殊效果:差异是真实的吗

现在你已经知道如何检验两个不同的群体或者两个独立群体的均值差异。很好,但这还不是全部。

如果你已经知道群体之间具有显著性差异,但是决定是否投入 64 000 美元的不仅仅是差异是否显著,还有是否有意义的问题。我们的问题是代表不同群体的数据分布之间的差异是否足以说明你观察到的差异,以及你检验到的差异就是真实的差异!好吧,现在进入效应量的讨论。

效应量(effect size)是对两个群体之间如何相互不同的测量——也是对处理规模的测量。类似于多大才算大。而且计算效应量特别有趣的方面是不考虑样本规模。计算效应量并且对效应量做出判断是理解显著性结果的全新的维度。

现在看下面的案例。研究者检验的问题是参与社区自助服务(例如扑克牌游戏、野外旅行等)是否提高了美国老年人的生活质量(从 1 到 10 分为十个等级)。研究者执行了为期六个月的服务项目,在项目期结束之后测量两个群体的生活质量(每个群体由 50 名 80 岁以上的老人构成,其中一个群体得到服务而另一个没有得到)。下面是结果:

	没有社区服务	有社区服务
均值	7.46	6.90
标准差	1.03	1.53

而且结论是在风险水平 0.034 下差异是显著的(也就是 $p<.05$)。

现在,差异是显著的,但是差异的规模是多大?

雅各布·科恩(Jacob Cohen)是对效应量贡献最大的人,他撰写了有关效应量最有影响、最重要的文章。他撰写的一本非常重要非常有影响的书(你的统计学老师的书架上肯定有这本书),指导研究者就差异和变量之间的关系提出的不同问题计算效应量。下面是具体介绍。

计算和理解效应量

与其他统计技术一样,计算效应量的方式也有多种。我们现在介绍最简单、最直接的方法。你会在我们下面给出的一些文献中对效应量有更多的了解。

到目前为止,最直接、最简单的计算效应量的方式就是以均值之间的差除以任何一个群体的标准差。这样做有一定的风险——因为假定两个群体的标准差(和方差)相等。就我们上面的案例来说,我们会这样计算:

$$ES = \frac{\overline{X}_1 - \overline{X}_2}{S} \tag{11.6}$$

其中:ES 表示效应量,

$\overline{X}_1$ 表示群体 1 的均值,

$\overline{X}_2$ 表述群体 2 的均值,

S 表示任何一个群体的标准差。

因此,在我们的案例中,

$$ES = \frac{7.46 - 6.90}{1.53} = 0.366 \tag{11.7}$$

结果等于 0.366,也就是这个案例的效应量是 0.37。

这意味着什么?科恩等人指出非常重要的一点是效应量的大、中、小的范围。他们使用下面的标准:

- 小效应量的范围是 0~0.2。
- 中等效应量的范围是 0.2~0.5。
- 大效应量的范围是 0.5 及以上。

我们案例的效应量是 0.37,属于中等规模。但是这实际上意味着什么呢?

效应量给我们这样的认识,即每个群体相对于另一个群体的位置。例如,如果效应量为 0,就意味着两个群体非常相似而且几乎完全重叠——两个数值分布之间没有差异。另一方面,效应量为 1 意味着两个群体大约有 45%(通常能达到这个量)重叠。而且,就如你所预期的,效应量越大就意味着两个群体重叠的部分越少。

雅各布·科恩的《行为科学的统计检定力分析》(*Statistical Power Analysis for the Be-*

havioral Sciences)[1967 年首次出版,最近的版本(1988)由 Lawrence Erlbaum Associates 公司出版],是每一个想了解更多信息的人所必需的。这本书有许多表格和技术,可以用于理解为什么统计显著的发现只是分析工作的一半——另一半是效应的规模。



如果你真的想对效应量了解更多,你可以通过简单的方式计算效应量,就如我们刚才所展示的(均值相减,然后除以任何一个标准差),或者偷看坐在你旁边的学霸同学。效应量计算公式的发展就是在上面的 *ES* 等式的分母中使用综合的方差。综合的标准差类似于群体 1 的标准差和群体 2 的标准差的平均值。公式如下:

$$ES = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}} \tag{11.8}$$

其中:*ES* 表示效应量,

$\overline{X}_1$ 表示群体 1 的均值,

$\overline{X}_2$ 表示群体 2 的均值,

σ_1^2 表示群体 1 的方差,

σ_2^2 表示群体 2 的方差。

如果我们将上面给出的数字代入这个公式,就会得到效应量为0.43,与我们使用之前给出的更简便的方法计算所得的 0.37 差异很大(但是都属于中等规模效应)。但是这是更准确的计算方法,大家应该好好地了解。

非常好用的效应量计算器

统计学家李·贝克尔(Lee Becker)开发了一个效应量计算器。利用这个计算器,你只要输入数值,单击 Compute,软件就会完成接下来的工作,如图 11.7 所示。谢谢贝克尔博士!

Group 1		Group 2	
M_1	7.4	M_2	6.9
SD_1	1.03	SD_2	1.53
Compute		Reset	
Cohen's *d*	0.38336	effect-size *r*	0.18826

小 结

t 检验是进行真实的统计检验,并从应用的角度完整地理解显著性的第一步。在进一步深入学习之前要确定你理解本章的内容。而且可以手动完成我们要求做到的一些计算。接下来我们学习相同检验的另外一种形式,不过这项检验是对同一个参与者群体进行两次测量而不是对不同的两个群体各进行一次测量。

练　习

1.使用名称为第 11 章数据集 1 的数据文件，在 0.05 的显著水平下检验研究假设，也就是“在课堂上男生比女生更经常举手”。使用计算器手动完成这次练习。那么就研究假设你得出的结论是什么？记住首先要确定这是单侧还是双侧检验。

2.使用相同的数据集（第 11 章数据集 1），在 0.01 的显著水平下检验研究假设，也就是“在课堂上男生和女生举手的次数不同”。使用计算器手动完成这次练习。那么就研究假设你得出的结论是什么？你使用的数据和问题 1 使用的数据相同，但是假设不同（一个是有方向的，另一个是无方向的）。那么结果会有何不同？为什么？

3.使用名称为第 11 章数据集 2 的数据文件，检验零假设，也就是“农村居民和城市居民对枪支管制有相同态度”。使用 T.TEST 函数检验零假设。

4.周五下午你要向老板汇报工作，你想要让她知道特拉华州纽瓦克市的两个商店每周出货率是相同还是不同。使用名称为第 11 章数据集 3 的数据文件和分析工具库来让她明白。最好现在就开始分析。

5.如果差异在统计上是显著的，但是效应量没有意义，这怎么解释？

(又是)两个群体的 t 检验——两个相关群体的均值检验

12

本章你会学到什么 ☺☺☺ ☒☒

- 何时使用非独立均值的 t 检验
- 如何计算观察值 t 值
- 解释 t 值,理解 t 值的含义
- 如何使用 T.TEST 函数和 T.DIST 函数
- 如何使用 Excel 中工具库中的 t-test 工具计算 t 值

非独立样本 t 检验介绍

如何给孩子提供最好的教育,很明显是任何社会都面对的重要问题之一。孩子之间差异太大,所以不得不在满足整体的基本需求和保证特殊的孩子(差异连续体的两端)得到他们需要的机会之间寻求平衡。阅读显然是教育过程中重要的部分,阿拉巴马大学的三位教授研究了阅读资料和常规教学对有学习障碍的孩子的阅读能力的影响。雷尼塔·古德曼、加里·萨普与安·舒梅特·福斯特(Renitta Goldman, Gary L. Sapp, & Ann Shumate Foster)发现,大体上来说,对阅读资料和课堂教学给予一年的日常指导在阅读成绩提高方面没有差异。通过对阅读资料这一组前测试和后测试的特定比较,他们发现 $t_{(34)} = 1.23$, $p>0.05$。在项目的最初,阅读资料组的孩子的阅读成绩是 85.5。在项目的最后,阅读资料组的孩子的阅读成绩是88.5——有差异,但是差异不显著。

为什么用非独立均值检验? 非独立均值检验表明是相同的群体在两种不同的条件下进行相同的研究。在这个案例中,条件是实验前和实验后。相同的一群孩子测试两次,也就是在 1 年期项目开始之前和 1 年期项目结束之后,所以我们使用非独立均值检验。依据上面的结果你可以知道,在项目前和项目后阅读成绩没有差异。t 值(1.23)非常小,没有落在我们拒绝零假设的值域范围之外。换句话说,变化太小,我们不能说变化是由随机因素之外的因素引起。这么小的差异——2.7(88.5-85.8)可能是由于抽样误差或者组内的变化引起。

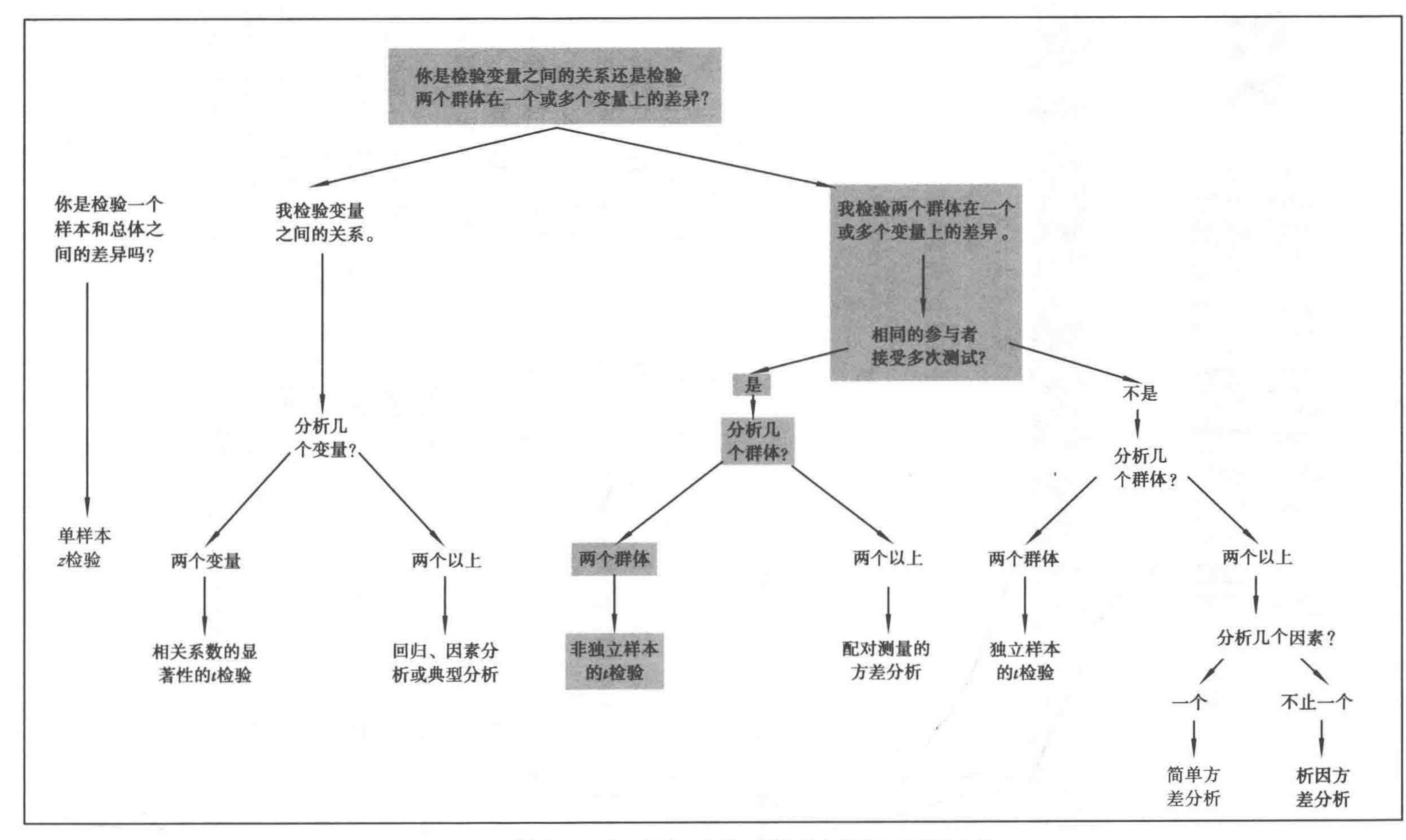


图12.1　确定非独立均值 t 检验是合适的统计检验方法

是否想了解更多？查阅古德曼、萨普和福斯特（Goldman, R., Sapp, G.L., & Foster, A.S.）1998 年发表在《感知与运动技能》（*Perceptual and Motor Skills*）86 期 192～194 页上的文章“学习障碍学生在阅读资料和常规课堂教学中阅读能力的提高（Reading achievement by learning disabled student in resource and regular classes.）”。

智慧和知识之路

下面介绍如何使用流程表选择合适的统计检验，也就是非独立均值的 t 检验。沿着图 11.1 中指示的步骤就可以。



1—学生的项目前测试和项目后测试成绩的差异是关注的重点。
2—每一个参与者接受不止一次的测试。
3—有两组数据。
4—合适的统计检验方法是非独立均值的 t 检验。



统计学家讨论非独立检验还有另外一种方式——重复测量。非独立检验也叫作重复测量，是在不同时间、不同条件、不同因素下进行，而且是对相同对象重复测量，相同对象可能是一个人、一个地点或者一个事物。

计算检验统计量

非独立均值 t 检验包含不同数据组均值的比较，而且重点是不同数值之间的差异。就如公式 12.1 所示，两次测试的差异总和构成分子，表示两组数据组之间的差异。

$$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - \left(\sum D\right)^2}{n-1}}} \tag{12.1}$$

其中：$\sum D$ 表示两组数据间差异的总和，

$\sum D^2$ 表示两组数据间差异的平方和，

n 表示成对观察的参与者数量。

下面用一些数据说明 t 值如何计算。就如上面给出的例子一样列出前测试和后测试结果，而且为了解释的需要假定这是开展阅读项目之前和之后的成绩。

下面是著名的八个步骤和 t 检验统计量的计算。

1.零假设和研究假设的表述。零假设表示前测试和后测试的阅读成绩均值之间没有差异。研究假设是单侧、有方向的假设，因为研究假设假定后测试成绩高于前测试成绩。

	前测试	后测试	差异	D^2
	3	7	4	16
	5	8	3	9
	4	6	2	4
	6	7	1	1
	5	8	3	9
	5	9	4	16
	4	6	2	4
	5	6	1	1
	3	7	4	16
	6	8	2	4
	7	8	1	1
	8	7	-1	1
	7	9	2	4
	6	10	4	16
	7	9	2	4
	8	9	1	1
	8	8	0	0
	9	8	-1	1
	9	4	-5	25
	8	4	-4	16
	7	5	-2	4
	7	6	-1	1
	6	9	3	9
	7	8	1	1
	8	12	4	16
总和	158	188	30	180
均值	6.32	7.52	1.2	7.2

零假设是：

$$H_0: \mu_{posttest} = \mu_{pretest} \tag{12.2}$$

研究假设是：

$$H_1: \overline{X}_{posttest} > \overline{X}_{pretest} \tag{12.3}$$

2.设置零假设的风险水平(显著性水平,或第一类错误)。风险水平或第一类错误或显著水平是 0.05,这完全由研究者决定。

3.选择合适的检验统计量。使用图 12.1 所示的流程图,我们确定合适的检验方法是非独立均值的 t 检验。因为这两个群体相互不独立,所以不是独立均值的 t 检验。实际上,这两个群体不是参与者群体而是相同参与者的两组成绩。两组之间相互依赖。非独立均值 t 检验的另一个名称是配对样本 t 检验,或相关样本 t 检验。你会在第 14 章看到,两组成绩(前测和后测)之间相关关系的显著性检验,和我们这里计算的 t 值之间有很大关系。

4.计算检验统计值(也叫作实际值)。现在我们代入观察值并进行计算。上面已经给出 t 值计算公式。代入具体的值之后,我们就得到等式 12.4(我们已经计算了前测和后测成绩的均值和标准差)。

$$t = \frac{30}{\sqrt{\dfrac{25 \times 180 - 30^2}{25 - 1}}} \tag{12.4}$$

代入具体的数值之后,我们得到下面的等式和最终的实际 t 值 2.45。项目前测试成绩

的均值是 6.32，项目后测试成绩的均值是 7.52。

$$t = \frac{30}{\sqrt{150}} = 2.45 \tag{12.5}$$

5.使用特定统计量的临界值分布表确定拒绝零假设需要的值。现在我们需要查阅附录 B 的表 B2，表 B2 列出了 t 检验的临界值。我们又一次进行了 t 检验，而且应用第 11 章中用到的相同的临界值表来确定拒绝零假设的临界值。

我们的第一个任务是确定自由度（df），自由度近似于样本规模。对现在选定的检验统计量来说，自由度是 $n-1$，其中 n 等于配对观察的参与者数量，也就是 $25-1=24$。这项统计检验的自由度是独特的，对其他统计检验来说并不需要。

使用这个数字（24）、你愿意承担的风险水平（之前定义的0.05）以及单侧检验（因为研究假设有方向——后测成绩大于前测成绩），拒绝零假设需要的值是 1.711。

6.比较实际值和临界值。实际值是 2.45，大于拒绝零假设值需要的临界值。

7.和 8.做出决定。

如果实际值大于临界值就不能接受零假设。如果实际值没有超过临界值，零假设就是最有力的解释。在这个案例中，实际值超过临界值——这个值足够大，我们可以说前测成绩和后测成绩的差异的确是由于随机因素之外的因素引起的。如果我们的实验安排正确，那么是什么因素影响结果？很简单——是日常阅读项目的引入。我们知道差异的产生是由于特定的因素。前测群体和后测群体之间的差异不可能是由随机因素引起的，而是由于特定的处理因素。

如何解释 $t_{(24)}=2.45, p<0.05$

- t 表示我们所用的检验统计量。
- 24 是自由度数值。
- 2.45 是实际值，是使用本章之前给出的公式计算所得的值。
- $p<0.05$（p 值是最重要的）表示对零假设的任何检验来说，后测成绩的均值大于前测成绩的均值是缘于随机因素的可能性小于 5%，也就是说有其他因素在发挥作用。因为我们以 0.05 作为研究假设比零假设更有解释力的标准，我们的结论就是两组成绩之间具有显著性差异。也就是其他因素在发挥作用。

使用 Excel 的 T.TEST 函数

非常有意思的是，Excel 使用相同的函数计算 t 值来检验两个非独立群体之间的差异（就如在第 11 章是检验两个独立群体之间的差异）。T.TEST 函数返回的结果是实际值出现概率，你几乎可以按照我们在第 11 章所展示的方式使用这个函数（你只要改变一个选项即可，在下面你就会看到）。

更多 Excel

除了 T.TEST 函数，还有 T.DIST 函数。只要你输入 t 值、自由度，假设是单侧还是双侧检验，Excel 就可以返回所得结果的概率。举个例子，你还记得第 8 章正态曲线的内容吧？如果 t 值是 1.96，自由度是 10 000，对应的概率值就是 0.05。这下你就清楚了嘛！

在工作表中将具体的数值输入两列，两列的标签分别是 Pretest（前测数值）和 Posttest（后侧数值），如图 12.2 所示（我们使用的之前案例中使用的数据）。

选择你想输入 T.TEST 函数的单元格。在这个案例中，我们打算将 T.TEST 函数值返回到单元格 D1（选择这个位置没有任何特殊的原因）。

	A	B
1	Pretest	Postest
2	3	7
3	5	8
4	4	6
5	6	7
6	5	8
7	5	9
8	4	6
9	5	6
10	3	7
11	6	8
12	7	8
13	8	7
14	7	9
15	6	10
16	7	9
17	8	9
18	8	8
19	9	8
20	9	4
21	8	4
22	7	5
23	7	6
24	6	9
25	7	8
26	8	12

图 12.2 非独立均值 T.TEST 函数所用数据

D1 fx =T.TEST(A2:A26,B2:B26,1,1)

	A	B	C	D	E	F
1	Pretest	Postest		0.010991		
2	3	7				
3	5	8				
4	4	6				
5	6	7				
6	5	8				
7	5	9				
8	4	6				
9	5	6				
10	3	7				
11	6	8				
12	7	8				
13	8	7				
14	7	9				
15	6	10				
16	7	9				
17	8	9				
18	8	8				
19	9	8				
20	9	4				
21	8	4				
22	7	5				
23	7	6				
24	6	9				
25	7	8				
26	8	12				

图 12.3 使用 T.TEST 计算 t 值的概率

现在使用公式→其他函数→统计→T.TEST 菜单选项，或者使用我们在 1a 中讨论的“插入一个函数”的技术在单元格 D5 输入 T.TEST 函数。函数形式如下。

=T.TEST(array1,array2,tails,type),

其中：

array1 是第一个数据集的单元格地址（在这个案例中是 A2:A26）；

array2 是第二个数据集的单元格地址（在这个案例中是 B2:B26）；

tails 选 1 还是 2 取决于是单侧还是双侧检验。单侧（有方向的）是 1，我们现在这个案例是 1，双侧（没有方向）是 2；

type 是 1 还是 2 取决于方差是相等（在我们的案例中是 1）还是不相等（2）。

在这个案例中，图 12.3 所示的是完成的 T.TEST 函数表达式（注意：因为观察对象是配对的或者相互依赖的，所以 type 等于 1）。

=T.TEST(A2:A26,B2:B26,1,1)

单击确定，你就会看到返回的结果是 0.010 991。

就如我早前得出的结论一样，处理因素具有显著的影响，因为由于随机因素导致的前测和后测之间差异的概率小于 0.010 991 或者 1%。这几乎不可能发生。



再说一次，要记住关于 T.TEST 函数两个重要的方面。

- T.TEST 函数的的确确不计算 t 值。（想一想，这是很容易就犯的错误。）
- T.TEST 返回的是由于随机因素导致的 t 值发生的可能性。例如，对发生概率是.88（要记住，这个值最高就是 1 或者 100%）的 t 值，零假设的解释力很高（即两者无差异）。

使用奇妙的分析工具计算 t 值

我们会再一次发现，对于非独立均值差异的 t 检验来说，分析工具库可以给我们工具和所有的信息，我们需要这些信息来进行 t 值及其显著性的判断。现在就按照我们在 11 章操作的类似程序来进行计算，只是这次我们使用相同的数据并且选择 t-检验：平均值的成对二样本分析选项。

1.单击［数据］→［数据分析］（Data Analysis），你将看到如图 12.4 所示的对话框。

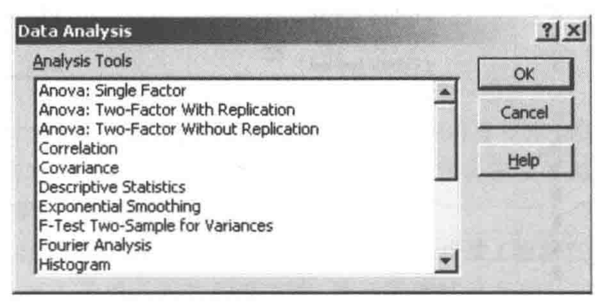


图 12.4 开始的分析工具库对话框

2.单击［t-检验：平均值的成对二样本分析］（t-Test: Paired Two Sample for Means）（这是 Excel 描述非独立均值的方式）选项，接着单击确定。你会看到图 12.5 所示对话框。

图 12.5 成对样本或者非独立均值检验

3.在［变量 1 的区域］（Variable 1 Range）输入前测数据的单元格地址。在图 12.2 所示的样本数据表中，单元格地址是 A1：A26（包括了数据列标签 Pretest）。

4.在［变量 2 的区域］（Variable 2 Range）输入后测数据的单元格地址。在图 12.2 所示的样本数据表中，单元格地址是 B1：B26（包括了数据列标签 Postest）。

5.单击［标志］（Labels）选项框，这样标签就可以包括进来。

6.单击[输出区域](Output Range)选项，输入你想让输出结果返回的单元格地址(同一工作表)。在这个案例中，我们将结果放置在以单元格 D1 开始的位置。

7.单击确定，就会看到如图 12.6 所示的结果，你得到很整齐(我做了些调整)的与分析相关的重要数据的汇总。在下表中列出了每一个统计量和相应的对其含义的描述。

	A	B	C	D	E	F
1	Pretest	Postest		t-Test: Paired Two Sample for Means		
2	3	7				
3	5	8			Pretest	Postest
4	4	6		Mean	6.32	7.52
5	6	7		Variance	2.98	3.34
6	5	8		Observations	25	25
7	5	9		Pearson Correlation	0.05	
8	4	6		Hypothesized Mean Difference	0	
9	5	6		df	24	
10	3	7		t Stat	-2.45	
11	6	8		P(T<=t) one-tail	0.01	
12	7	8		t Critical one-tail	1.71	
13	8	7		P(T<=t) two-tail	0.02	
14	7	9		t Critical two-tail	2.06	
15	6	10				
16	7	9				
17	8	9				
18	8	8				
19	9	8				
20	9	4				
21	8	4				
22	7	5				
23	7	6				
24	6	9				
25	7	8				
26	8	12				

图 12.6　成对样本 t 检验或者非独立均值 t 检验的分析工具库分析结果

再一次，我们之前的结论得到了支持。单侧检验的 t 值 2.45，由于随机因素导致的概率是 0.01，发生的可能性很小。我们的结论是什么？一定有其他因素在发生作用，这个其他因素就是一个干预项目的执行。

统计量	描述
Mean	每个变量的平均数值
Variance	每个变量的方差
Observations	每个数据组观察值的个数，因为观察对象是配对的——每个对象有两个观察值，因此两组的观察值数量始终是相等的
Pearson Correlation	两个变量之间关系的测量(回顾第 5 章学习的内容)
Hypothesized Mean Difference	你已经表明的你所预期的差异(返回到对话框)
df	自由度
t Stat	t 统计量的值
P(T<=t) one-tail	单侧检验，随机因素导致的 t 值的概率
t Critical one-tail	单侧检验，你需要超过的临界值(回顾第 9 章临界值内容)
P(T<=t) two-tail	双侧检验，随机因素导致的 t 值的概率
t Critical two-tail	双侧检验，你需要超过的临界值(回顾第 9 章临界值内容)



图 12.6 中 t 值为什么是负号？唯一的原因就是分析工具库计算 t 值的方法。计算方式始终是用第一列数据减去第二列数据，如果第二列数据更大，那么结果就是负值。如果你将后测数据作为第一列，那么-2.45 就会是 2.45，但是其他的值不会发生变化。重要的是 t 值是显著的（在这个案例中，对单侧检验来说，概率是 0.01）。

小　结

这是对均值的检验。你已经知道如何比较来自独立群体（第 11 章）和非独立群体（第 12 章）的数据，现在是时候更进一步学习处理两个以上群体（可能是独立的也可能是非独立的）的显著性检验。这项统计技术就是方差分析，这项统计分析工具非常重要、应用广泛，也很有价值。

练　习

1.独立均值检验和非独立均值检验有何不同？何时使用哪一个？

2.使用分析工具库计算第 12 章数据集 1 的 t 值，然后将结论写下来，也就是 25 个街区的废纸回收项目是否影响纸的使用量。（提示：实施项目前和实施项目后是两个不同的测试水平。）在 0.01 显著水平下检验零假设。

3.在下面的案例中，你要进行独立均值 t 检验还是非独立均值 t 检验？

a.脚腕扭伤的两组病人采用不同的治疗方式，哪一种治疗最有效？

b.护理研究者想知道得到额外家庭护理的病人的恢复是否快于只得到标准家庭护理的病人。

c.一组青少年得到了人际技能指导，接着在 9 月份和 5 月份接受测试来看这是否有利于他们家庭关系和谐。

d.一组成年男子得到降低高血压的指导，另一组没有得到任何指导，在这前后测量了他们的血压。

e.一组男子加入了一个为期 6 个月的锻炼项目。测试项目前后两个时期他们的心脏健康状况。

4.使用第 12 章数据集 2 手动计算 t 值，然后将结论写下来，也就是引入服务项目之后，使用服务中心的家庭的满意水平是否有差异。使用 Excel 完成这个练习，并且给出结果对应的准确概率。

5.二手车销售经理杰克不知道他支付 100 000 美元让他的二手车销售员接受培训是否有效果。他向优秀销售员戴维询问意见，戴维认为会起作用但是不知道作用多大。因此杰克雇佣了统计专家来分析培训是否产生了销售差异这个问题。下面是数据，你来告诉杰克培训是否发挥了作用。

月份	培训前	培训后
12 月	$ 58 676	$ 87 890
1 月	$ 46 567	$ 87 876
2 月	$ 87 656	$ 56 768
3 月	$ 65 431	$ 98 980
4 月	$ 56 543	$ 98 784
5 月	$ 45 456	$ 65 414
6 月	$ 67 656	$ 99 878
7 月	$ 78 887	$ 67 578
8 月	$ 65 454	$ 76 599
9 月	$ 56 554	$ 88 767
10 月	$ 58 876	$ 78 778
11 月	$ 54 433	$ 98 898

6.使用第 12 章数据集 3 分析，在入学时给予青少年一定的咨询是否影响他们对其他不同种族的青少年的宽容度。在咨询前和 6 个月后进行了影响评价。这个咨询项目是否发挥作用？结果变量是对他人态度的测量的得分，可能的分值是从 0 到 50，值越高宽容度越高。可选用任何方法来完成这项分析。

两个群体是否太多？——尝试进行方差分析 13

本章你会学到什么 ☺☺☺ ☒☒☒☒

- 方差分析是什么，何时使用
- 如何计算和解释 *F* 统计量
- 如何使用 F.TEST 和 F.DIST 函数
- 如何使用 Excel 中工具库中的单因素方差分析工具计算 *F* 值

方差分析介绍

心理学的一个新兴领域是体育心理学。虽然这个领域主要关注提高运动成绩，但也关注运动的许多其他方面。其中之一是什么心理技能对于成为一个成功的运动员是必需的。带着这个问题，马里斯·古塔、杨尼斯·塞奥佐拉基斯和乔治斯·卡拉莫萨利蒂斯（Marious Goudas, Yiannis Theodorakis, & Georgios Karamousalidis）对运动员应对技能量表的有效性进行了检验。

作为研究的一部分，他们使用简单的**方差分析**（analysis of variance 或 ANOVA）检验假设，也就是某项运动的训练年数与应对技能（或者运动员应对技能量表的得分）的关系。因为需要检验两个以上的群体，并比较这些群体的平均成绩，所以使用方差分析。具体地说，群体 1 是训练年数在 6 年之内的运动员，群体 2 是训练了 7~10 年的运动员，群体 3 是训练年数在 10 年以上的运动员。

方差分析的检验统计量是 *F* 统计量（以这个统计量的建立者 R.A.Fisher 命名），结果是 $F_{(2,110)}=13.08, p<0.01$。三个群体的压力下巅峰表现子量表测试成绩的均值两两不同。换句话说，测试成绩的不同是由于在运动方面的训练年数的不同，而不是可能影响成绩的随机因素。

是否想了解更多？查阅原始文献：古塔、塞奥佐拉基斯和卡拉莫萨利蒂斯（Goudas, M., Theodorakis, Y., & Karamousalidis, G.）1998 年发表在《感知和运动技能》（*Perceptual and Motor Skills*）86 期 59~65 页的文章"篮球运动中的心理技能：发展运动员应对技能量表的希腊形式的初步研究（Psychological skills in basketball: Preliminary study for development of a Greek form of the Athletic Coping Skills Inventory.）。"

智慧和知识之路

下面介绍如何使用图 13.1 所示流程图选择方差分析作为合适的统计检验方法。沿着

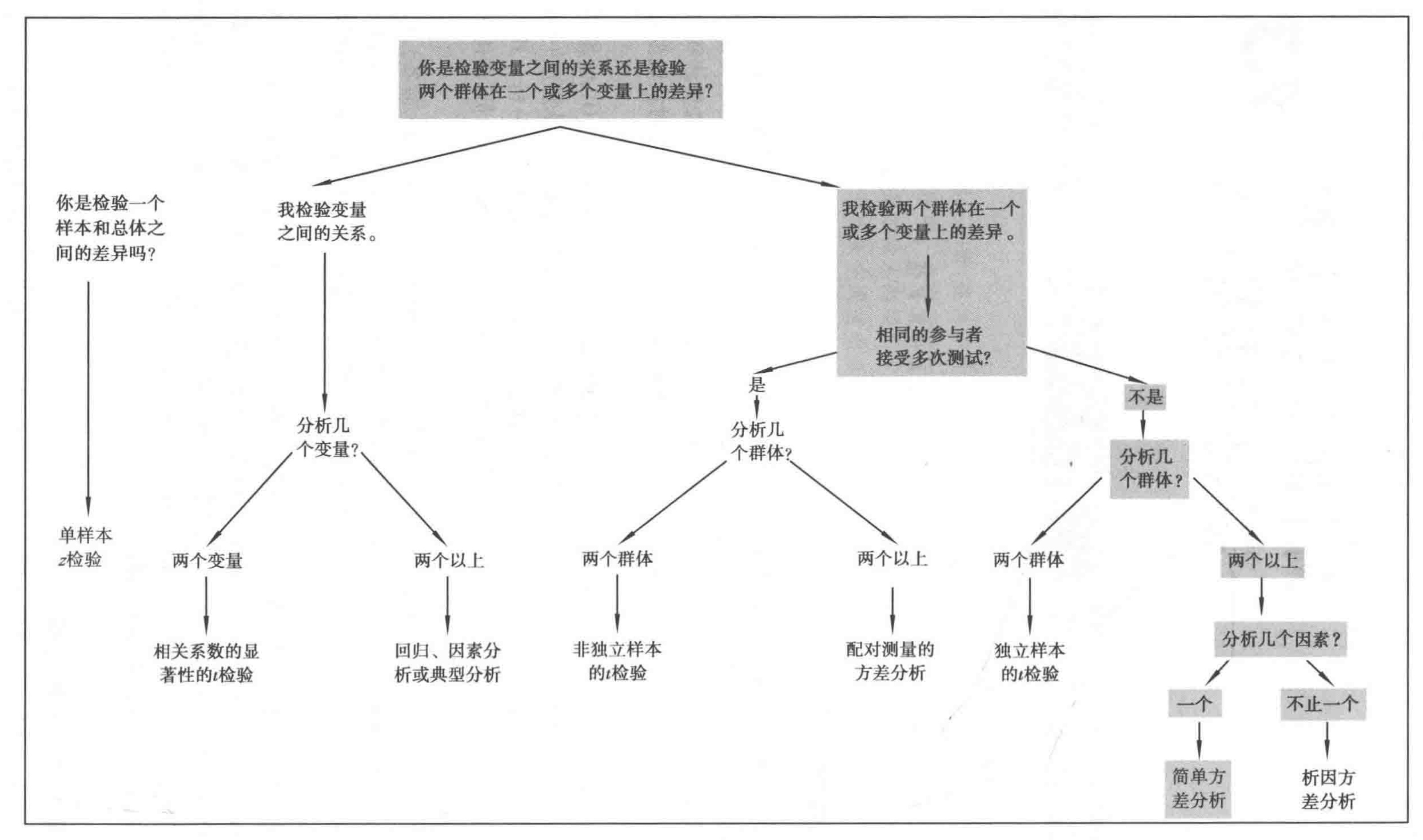


图13.1　确定方差分析是合适的统计检验方法

图中指示的步骤就可以。



1—我们检验不同群体之间的差异，在这个案例中是检验不同运动员群体巅峰成绩的差异。

2—每一个运动员只接受一次测试。

3—有三个群体（按训练年数分为6年以下、7~10年和10年以上）。

4—合适的检验统计是简单方差分析。

方差分析的不同类型

方差分析有许多不同的形式。最简单的形式是**简单方差分析**（simple analysis of variance），也是本章的重点，只分析一个因素或者一个处理变量（如群体身份），且有两个以上的群体受到这个因素的影响。简单方差分析也叫作**一元方差分析**（one-way analysis of variance 或 single factor），因为只有一个分组维度。这项技术叫作方差分析，这是由于运动成绩差异的方差可以分解为群体内个体差异产生的方差和群体之间差异产生的方差。接着对两类方差进行相互比较。

实际上，方差分析在许多情况下类似于 t 检验。在这两项检验中都需要计算均值之间的差异。但方差分析要处理两个以上的均值。

例如，我们调查每个星期待在预备学校5、10和20个小时对孩子语言发展的影响。每个孩子所属的组别就是处理变量，或者就是分组因素。语言发展是被解释变量或者是结果。实验设计类似如下所示：

组1（每星期5小时）	组2（每星期10小时）	组3（每星期20小时）
语言发展测试成绩	语言发展测试成绩	语言发展测试成绩

更复杂的方差分析类型是**析因设计**（factorial design），是分析一个以上的处理变量。下面的案例是研究参加预备学校的时间所产生的效应，但是性别差异产生的效应也是研究的内容。实验设计类似如下所示：

		待在预备学校的时间		
		组1（每星期5小时）	组2（每星期10小时）	组3（每星期20小时）
性别	男	语言发展测试成绩	语言发展测试成绩	语言发展测试成绩
	女	语言发展测试成绩	语言发展测试成绩	语言发展测试成绩

这是3×2的析因设计。3表示分组因素有三个层级（组1、组2和组3）。2表示其他分组因素有两个层级（男和女）。综合起来就有6种不同的可能性（每个星期待在预备学校5小时的男生，每个星期待在预备学校5小时的女生，每个星期待在预备学校10小时的男生，等等）。

析因设计遵循和简单方差分析一样的基本逻辑和原则，但是析因设计更为复杂，需要同时检验一个以上的因素的影响以及多因素综合的影响。别担心——接下来的一章会全

面学习析因设计。

计算 F 检验统计量

简单方差分析检验两个以上的群体在一个因素或一个维度上的均值差异。例如，你可能想知道四个群体（20、25、30 和 35 岁的年龄群体）对私人学校的公共财政支持的态度。或者你想知道 5 个不同年级（2、4、6、8 和 10 年级）的学生的父母群体参与学校活动的水平是否有差异。

任何分析如果

- 只有一个维度或者一个处理变量，
- 分组因素有两个以上的层级，
- 而且关注不同群体在平均成绩上的差异。

就需要使用简单方差分析。

F 统计量是检验"群体之间有差异"假设的检验统计量，计算公式如公式 13.1 所示。就方差分析而言这是简单的公式，但是相对于前几章学习的其他检验统计量来说需要花费更多精力来计算。

$$F = \frac{MeanSquares_{Between}}{MeanSquares_{Within}} \tag{13.1}$$

Tech Talk 这个比率背后的逻辑就是这样。如果组内完全没有变化（所有的成绩都相同），那么组之间的任何差异都有意义，对吧？可能如此。方差分析公式是一个比率，比较组间的变化量（由于分组因素产生）与组内的变化量（由于随机因素产生）。如果比值为 1，那么组内差异产生的变化量等于组间差异产生的变化量，而且组间的任何差异都不显著。如果组间差异的平均值变大（也就是比率的分子变大），F 值也变大。F 值越大，在所有 F 值的分布中就会越趋向于极值，也就越可能是随机因素之外的因素在产生影响。

下面给出一些数据和初步的计算来说明 F 统计量如何计算。就我们的例子来说，我们假定这是三个预备学校参与者群体和他们的语言测试成绩。

群体 1 的成绩	群体 2 的成绩	群体 3 的成绩
87	87	89
86	85	91
76	99	96
56	85	87
78	79	89
98	81	90
77	82	89
66	78	96
75	85	96
67	91	93

下面是著名的八个步骤和 F 检验统计量的计算。

1.零假设和研究假设的表述。公式 13.2 所示的零假设表示三个不同群体的均值之间没有差异。方差分析，也叫作 F 检验（因为计算得出的是 F 统计量或 F 比值），寻求不同群体之间所有的差异。

F 检验不是分析配对差异，如群体 1 和群体 2 之间的差异。我们需要使用另一项统计技术进行配对差异分析，本章后面部分会讨论。

$$H_0: \mu_1 = \mu_2 = \mu_3 \tag{13.2}$$

公式 13.3 所示的研究假设表示三个群体的均值之间相互有差异。要注意的是所有的差异之间没有方向，这是因为所有的 F 值都是无方向的。

$$H_1: \bar{X}_1 \neq \bar{X}_2 \neq \bar{X}_3 \tag{13.3}$$



前面我们已经讨论了单侧和双侧检验。在讨论方差分析时不需要确定单侧或双侧。因为要检验两个以上的群体，而且因为 F 检验是综合的检验（也就是检验均值之间的所有差异），讨论特定差异的方向没有意义。

2.设置零假设的风险水平（显著性水平或第一类错误）。风险水平是 0.05。再一次申明，风险水平完全由研究者决定。

3.选择合适的检验统计量。使用图 13.1 所示的流程图，我们确定合适的检验方法是简单方差分析。

4.计算检验统计值（也叫作实际值）。现在我们代入具体的值并进行计算。需要进行很多计算。

a.F 值是组间差异和组内差异的比值。要计算这些值，首先我们要计算每一种差异——组间、组内和总的差异的平方和。

b.组间差异平方和等于对所有值的均值和每一个群体的均值之差平方然后求和。这意味着每一群体的均值和总的均值的差异的大小。

c.组内差异平方和等于对群体内每一个具体的值和这个群体的均值之间的差异平方然后求和。这意味着群体内每一个值和这个群体的均值的差异的大小。

d.总的差异平方和等于组间差异平方和与组内差异平方和的总和。

现在我们计算这些值。依据上面给出的实际数值计算所有需要计算的组间、组内和总的平方和。

首先，我们看看这个扩展的表中所有的项目，先从图 13.2 的左下角开始：

n 表示每个群体的参与者数量（如 10），

$\sum X$ 表示每个群体的具体数值的总和（如 766），

$\bar{X}$ 表示每个群体的均值（如 76.60），

$\sum (X^2)$ 表示每个数值的平方和（如 59 964），

$\left(\sum X\right)^2/n$ 表示每个群体的所有数值的和的平方除以群体的规模（如58 675.60）。

其次，我们看看表的右下角的项目：

N 表示参与者的总体数量（如 30），

	群 体	考试成绩	X^2	群 体	考试成绩	X^2	群 体	考试成绩	X^2	
	1	87	7 569	2	87	7 569	3	89	7 921	
	1	86	7 396	2	85	7 225	3	91	8 281	
	1	76	5 776	2	99	9 801	3	96	9 216	
	1	56	3 136	2	85	7 225	3	87	7 569	
	1	78	6 084	2	79	6 241	3	89	7 921	
	1	98	9 604	2	81	6 561	3	90	8 100	
	1	77	5 929	2	82	6 724	3	89	7 921	
	1	66	4 356	2	78	6 084	3	96	9 216	
	1	75	5 625	2	85	7 225	3	96	9 216	
	1	67	4 489	2	91	8 281	3	93	8 649	
n	10			10			10			N=30.00
$\sum X$		766			852			916		$\sum X$=2 534.00
X		76.60			85.20			91.60		$(\sum\sum X)^2/N$=**214 038.53**
$\sum(X^2)$		59 964			72 936			84 010		$\sum\sum(X^2)$=**216 910**
$(\sum X)^2/n$		58 675.60			72 590.40			83 905.60		$(\sum X)^2/n$=**215 171.60**

图 13.2　计算一元方差分析需要的重要的值

$\sum\sum X$ 表示所有群体的数值的总和，

$\left(\sum\sum X\right)^2/N$ 表示所有数值的总和的平方除以 N，

$\sum\sum\left(X^2\right)$ 表示所有数值的平方的总和，

$\sum\left(\sum X\right)^2/n$ 表示每个群体和的平方的总和除以 n。

以上是所有需要进行的计算，我们也几乎完成了计算。

首先我们计算所有不同来源的差异的平方和，也就是进行这样的计算：

组间平方和	$\sum\left(\sum X\right)^2/n-\left(\sum\sum X\right)^2/N$，或者 215 171.60−214 038.53	1 133.07
组内平方和	$\sum\sum\left(X^2\right)-\sum\left(\sum X\right)^2/n$，或者 216 910 − 215 171.6	1 738.40
总平方和	$\sum\sum\left(X^2\right)-\left(\sum\sum X\right)^2/N$，或者 216 910 − 214 038.53	2 871.47

其次，我们需要计算平方和的均值，也就是简单的平方和的平均值。这些都是最终计算 F 比率需要的方差估计。

我们是以近似的自由度（*df*）去除每一个平方和。还记得吗？自由度是样本规模或者群体规模的近似值。对方差分析来说我们需要两类自由度。对组间估计来说，自由度是 $k-1$，其中 k 等于群体的数量（在这个案例中，有 3 个群体，自由度是 2）；对组内估计来说，我们需要的自由度是 $N-k$，其中 N 是总的样本规模（也就是说自由度是 30−3，或 27）。而且 F 比率是组间差异平方和均值与组内差异平方和均值的简单比值，即566.54/64.39=8.799。这就是实际的 F 值。

下面的表格总结了用于计算 F 比率的方差估计，也是专业期刊和手稿中大多数 F 表出现的形式。

来源	平方和	*df*	平方和的均值（均方）	*F*
组间	1 133.07	2	566.54	8.799
组内	1 738.40	27	64.39	
总和	2 871.47	29		

要计算一个小小的 F 值确实挺麻烦，但是我们之前已经说过，至少手动计算一次对了解计算过程很重要。计算过程会告诉你这些数字的来源，而且一定程度上可以更好地了解这些数字的含义。



你已经了解 t 检验，所以你可能想知道 t 值（总是用于两个群体的均值差异的检验）和 F 值（总是用于两个以上的群体）之间的关系。有趣的是两个群体的 F 值等于两个群体的 t 值的平方，即 $t^2=F$ 或 $t=\sqrt{F}$。很简单，但是只知道其中一个却还想知道另一个的情况下就非常有用。

5.使用特定统计量的临界值分布表确定拒绝零假设需要的值。就如之前所进行的，我们需要比较实际值和临界值。我们现在需要查阅附录B表B3，也就是F检验的临界值分布表。我们的第一个任务是确定分子的自由度，也就是$k-1$，或$3-1=2$。接着确定分母的自由度，也就是$N-k$，或$30-3=27$。结合在一起可以表示为$F_{(2,27)}$。

实际值是8.80，或$F_{(2,27)}=8.80$。在显著水平为0.05、分子自由度为2（由表B3中的纵行表示）、分母自由度为27（由表B3中的横行表示）情况下临界值是3.36。也就是在显著水平为0.05、自由度是2和27的三个群体的均值的综合检验来说，拒绝零假设所需要的值是3.36。

6.比较实际值和临界值。实际值是8.80，在0.05显著水平下拒绝零假设，也就是说三个群体相互有差异（没有关注差异在什么地方）的临界值是3.36。

7.和8.做出决定。

如果实际值大于临界值就不能接受零假设。如果实际值没有超过临界值，零假设就是最有力的解释。在这个案例中，实际值超过临界值——这个值足够大，我们可以说三个群体之间的两两差异不是由于随机因素引起的。如果我们的实验过程正确，那么是什么因素影响结果？很简单——是在预备学校的时间。我们知道差异的产生是由于特定的因素，因为群体之间的差异不可能是由随机因素引起的，而是缘于特定的处理变量。

如何解释 $F_{(2,27)}=8.80, p<0.05$

- F表示我们使用的检验统计量。
- 2，27是组间估计和组内估计的自由度数值。
- 8.80是实际值，是使用本章之前给出的公式计算所得的值。
- $p<0.05$（p值是最重要的）表示对零假设的任何检验来说，每个群体语言技能的平均成绩相互不同的原因是由于随机因素而不是实验变量的影响造成的可能性小于5%。因为我们以0.05作为研究假设比零假设更有解释力的标准，我们的结论就是三个群体之间存在显著差异。

Tech Talk

想象这样的情境。你是广告公司的有高级授权的研究者，想知道颜色是否影响销售。而且你在0.05的显著水平下进行检验。你将全黑、全白、25%带彩色、50%带彩色和100%是彩色的产品集合在一起构成5个不同的层级。进行方差分析并发现有差异存在。但是方差分析是综合的检验，你不知道显著差异的来源。因此你一次只能选择两个群体（如25%带彩色和75%带彩色）进行相互检验。实际上你要检验每一个两两群体组合的差异。这样做对吗？不对。这被称为多元t检验，而多元t检验实际上违反一些原则。当进行多元t检验时，由于要执行的检验的数量增加，第一类错误（你设定为0.05）发生的可能性提高。这项检验中有10对可能的群体差异比较（如没带彩色与25%带彩色，没带彩色与50%带彩色，没带彩色与75%带彩色等），第一类错误真正发生的可能性是$1-(1-a)^k$，其中：

α表示第一类错误发生的水平，在这个案例中是0.05；

k表示比较的数量。

因此每一对比较群体被检验时实际的第一类错误不是0.05，而是0.40，

$$1-(1-0.05)^{10}=0.40(!!!!!)$$

肯定不是 0.05。差异相当大,不是吗?

使用 Excel 的 F.DIST 和 F.TEST 函数

比较有趣的是,如 T.TEST 函数一样,Excel 没有计算 F 统计量具体值的函数。F.DIST 函数和 F.TEST 函数只计算两组数据的值(上一章我们使用 T.TEST 函数和 T.DIST 函数以及分析工具库)。因此,我们就直接使用分析工具库,它提供非常强大的 ANOVA 工具。

使用奇妙的分析工具计算 F 值

ANOVA 比本书中到现在所学的其他推论统计工具更复杂,ANOVA 分析包工具非常好,能体现通过使用 Excel 学习统计学的价值。

分析工具库中有三个方差分析选项:

- 方差分析:单因素方差分析
- 方差分析:可重复双因素方差分析
- 方差分析:无重复双因素方差分析

我们在本部分学习使用三个中的第一个选项,其他两项在下一章析因方差分析中学习。我们使用本章之前提供的数据,具体数据如图 13.3 所示,每一列代表不同的处理水平(每周 5、10、20 小时)。

(因为你们中的一些人等不及到下一章,就想了解其他两个选项,所以先简单解释一下,方差分析的"重复"部分是指同一对象被测试不止一次——有些类似于非独立均值 t 检验。)

	A	B	C
1	Group 1 Language Scores (5 Hours)	Group 2 Language Scores (10 hours)	Group 3 Language Scores (20 hours)
2	87	87	89
3	86	85	91
4	76	99	96
5	56	85	87
6	78	79	89
7	98	81	90
8	77	82	89
9	66	78	96
10	75	85	96
11	67	91	93

图 13.3 单因素方差分析数据

1.点击[数据]→[数据分析](Data Analysis),你就会看到如图 13.4 所示数据分析对话框。

2.点击[方差分析:单因素方差分析](Anova: Single Factor),接着点击[确定](OK),你就会看到图 13.5 所示的对话框。

3.在[输入区域](Input Range)选项中,输入三组数据的单元格地址。在图 13.3 所示的样本工作薄中,单元格地址是 A1:C11(包含了三组数据的标签)。

4.点击选择数据[分组方式](Grouped By):是按照行(Rows)或者列(Columns)分组。

5.点击[标志位于第一行](Label in First Row)这样可以将标签包括进来。

6.点击[输出区域](Output Range)选项,输入你想让结果出现的单元格位置(同一工

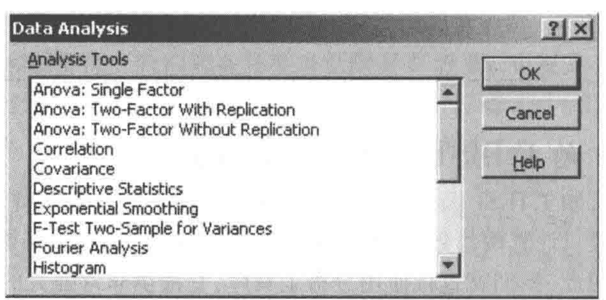


图 13.4　启动分析工具库的对话框

图 13.5　单因素方差分析对话框

作表)。在这个案例中,我们将结果放在以单元格 A13 起始的位置。

7.点击[确定](OK),你就会看到图 13.6,你会得到排列整齐(我们做了一些整理工作)的重要的与数据分析相关的统计量的汇总。在下表中列出了每个统计量并对其含义进行了描述。

	A	B	C	D	E	F	G
1	Group 1 Language Scores (5 Hours)	Group 2 Language Scores (10 hours)	Group 3 Language Scores (20 hours)				
2	87	87	89				
3	86	85	91				
4	76	99	96				
5	56	85	87				
6	78	79	89				
7	98	81	90				
8	77	82	89				
9	66	78	96				
10	75	85	96				
11	67	91	93				
12							
13	Anova: Single Factor						
14							
15	SUMMARY						
16	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
17	Group 1 Language Scores (5 Hours)	10	766	76.6	143.16		
18	Group 2 Language Scores (10 hours)	10	852	85.2	38.4		
19	Group 3 Language Scores (20 hours)	10	916	91.6	11.6		
20							
21							
22	ANOVA						
23	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
24	Between Groups	1133.07	2	566.53	8.80	0.00	3.35
25	Within Groups	1738.4	27	64.39			
26							
27	Total	2871.47	29				

图 13.6　单因素方差分析的输出结果

统计量	描 述
Groups	列出每一组
Count	每一组的观察数量
Sum	每一组的观察值的总和
Average	每一组的均值
Variance	每一组的方差
Source of Variation	误差的来源
SS	误差平方和
df	与误差来源对应的自由度
MS	均方(误差平方和均值)
F	*F* 值 (实际值)
P-value	*F* 值的显著性水平
Fcrit	拒绝零假设的 *F* 分布临界值

就如我们之前所讨论的,这里的 *F* 值已经大到我们不能认为差异只是由于偶然因素引起的。因此,我们的结论是在预备学校的时间不同会带来语言发展方面的差异。

小 结

方差分析(单因素方差分析或者析因方差分析)是这本书中你能学到的最复杂的推论检验。你需要付出更多的精力来进行手动计算,即使使用 Excel 时,你也必须了解这是综合的检验,而且不能给出配对群体差异的信息。我们还有均值检验的方法,析因方差分析——Holy Grail 方差分析,可以分析两个或更多因素。我们会在第 14 章学习。

练 习

1.使用下面的表分别给出简单一元方差分析的案例、两个两因素方差分析的案例和一个三因素方差分析的案例。我们给出一些范例,比照我们所做的确定分组变量和检验变量。

设计	分组变量	检验变量
简单方差分析	培训时间分为四个层级——2、4、6和8个小时	打字的准确程度
	填入你的案例	填入你的案例
	填入你的案例	填入你的案例
	填入你的案例	填入你的案例
两因素方差分析	培训的两个层级和性别(2×2设计)	
	填入你的案例	填入你的案例
	填入你的案例	填入你的案例
三因素方差分析	培训的三个层级、性别和收入的三个层级	选举态度
	填入你的案例	填入你的案例

2.使用第 13 章数据集 1 和 Excel，计算游泳者每周训练的平均时间（<15，15~25 以及>25 小时）的三个层级比较的 F 比率，被解释变量或结果变量是 100 码自由泳时间。回答这个问题，也就是训练时间是否产生差异。可以使用分析工具库或手动计算。

3.斯蒂芬认识到广告吸引注意有许多的技巧，他想检验样本产品的三种广告方式，也就是彩色的、黑白色或者色彩组合。下面是每个产品以 1~10 为测量分值的吸引力数据。现在他想知道三种形式是否存在吸引力差异。

彩色	黑与白	组合色
10	4	9
8	5	8
7	4	8
8	3	9
9	3	8
6	4	7
7	5	8
6	6	9
6	5	9
7	7	10
8	6	10
7	5	9
6	4	8
5	5	9
6	4	10
7	4	10
7	3	8

4.使用第 13 章数据集 2，研究者想了解 5 所高中 10 年间毕业率（以百分比表示）是否存在差异。（提示：年限是一个因素吗？）

两个因素——析因方差分析 14

本章你会学到什么 ☺ ☒☒

- 何时使用一个以上因素的方差分析
- 什么是主效应和交互效应
- 如何使用分析工具库完成析因方差分析

析因方差分析介绍

人们如何做出决策已经成为迷惑心理学者几十年的问题。相关研究形成的数据广泛地应用在广告学、商业、规划和宗教领域。米尔蒂亚德·普罗阿斯与乔治·多加尼斯(Miltiadis Proios, & George Doganis)研究积极参与决策过程(在一系列的条件下)的经验和年龄如何对道德归因产生影响。研究样本由 148 个裁判构成——56 个足球裁判、55 个篮球裁判以及 37 个手球裁判。他们的年龄范围是 17~50 岁,性别不是考虑的重要变量。在整个样本中,大约 8%没有社会、政治或运动领域全面参与决策制定过程的任何经验,大约 53%表现积极但是没有全面参与,大约 39%既表现积极,又全面参与一些组织中的决策制定。二元方差分析(在第 18 章可以更多了解多元方差分析)可以展现经验和年龄对道德归因和裁判的目标定位的交互影响。

什么是二元方差分析? 很容易——有两个独立因素,第一个是经验水平,第二个是年龄。与每一个方差分析程序相类似,包括:

- 对年龄的主效应的检验,
- 对经验的主效应的检验,
- 对经验和年龄的交互效应的检验(结果可能是显著的)。

在检验一个以上因素或者独立变量时就体现出方差分析的一个重要特点,也就是研究者可以分析每一个因素的效应,同时可以通过交互效应分析两者共同的效应,在本章后面的部分会进行更多的讨论。

是否想了解更多? 查阅原始文献:普罗阿斯与多加尼斯(Proios, M., & Doganis, G.) 2003 年发表在《感知与运动技能》(*Perceptual and Motor Skills*)第 96(1)期 113~126 页的文章"积极活动的经验、决策制定过程的参与和年龄对裁判的道德归因和目标定位的影响。(Experience from active membership and participation in decision-making process and age in moral reasoning and goal orientation of referees.)。"

析因方差分析的两种类型

析因方差分析有两种类型。Excel 中的分类一是“方差分析：可重复双因素方差分析”，二是“方差分析：无重复双因素方差分析”。两种类型中都有两个因素，差别是相同对象就某个因素测量的次数。

例如，无重复方差分析中我们可以测量两个因素的效应，这两个因素是居住地（城市或者农村）和选举背景（绿党或者非绿党）。我们使用对环境破坏的态度作为结果变量。在这个研究设计中（下面会展示），在四个单元格中都是不同的独立的观察对象。没有人最初的居住地可以既是城市又是农村，也没有人在一个选区可以既是绿党又是非绿党。这里没有重复。

		最初居住地	
		城市	农村
政治背景	绿党		
	非绿党		

另外一方面，在有重复的方差分析中，我们检验两个因素的效应，这两个因素是时间变化（从 9 月到 7 月）和学习科目（数学或者拼写）。我们使用成绩作为结果变量。在这个案例中，重复是指“时间变化”这个变量上的重复，因为相同的（这就是重复产生之处）科目被测试了两次。下面就是研究设计。

		时间	
		九月	七月
成绩	数学		
	拼写		

更多 Excel

你知道吗，Excel 实际上是人们编制的最酷的个人电脑应用软件。但是它也有自身的缺点。一个是方差分析工具，不论是否重复的方差分析，Excel 不能提供同样的清晰的输出结果。因为无重复方差分析的选项不能提供这项技术应用的清晰案例，我们不会纠结在让人苦恼的细节方面（这也是你们选择这本书的原因），只处理重复方差分析选项。可以看看我们在第 19 章提供的软件介绍，你可选择其他你想要使用的软件。

智慧和知识之路

下面介绍如何使用图 14.1 所示流程图选择方差分析（现在处理一个以上的因素）作为合适的检验统计。沿着图中指示的步骤就可以。

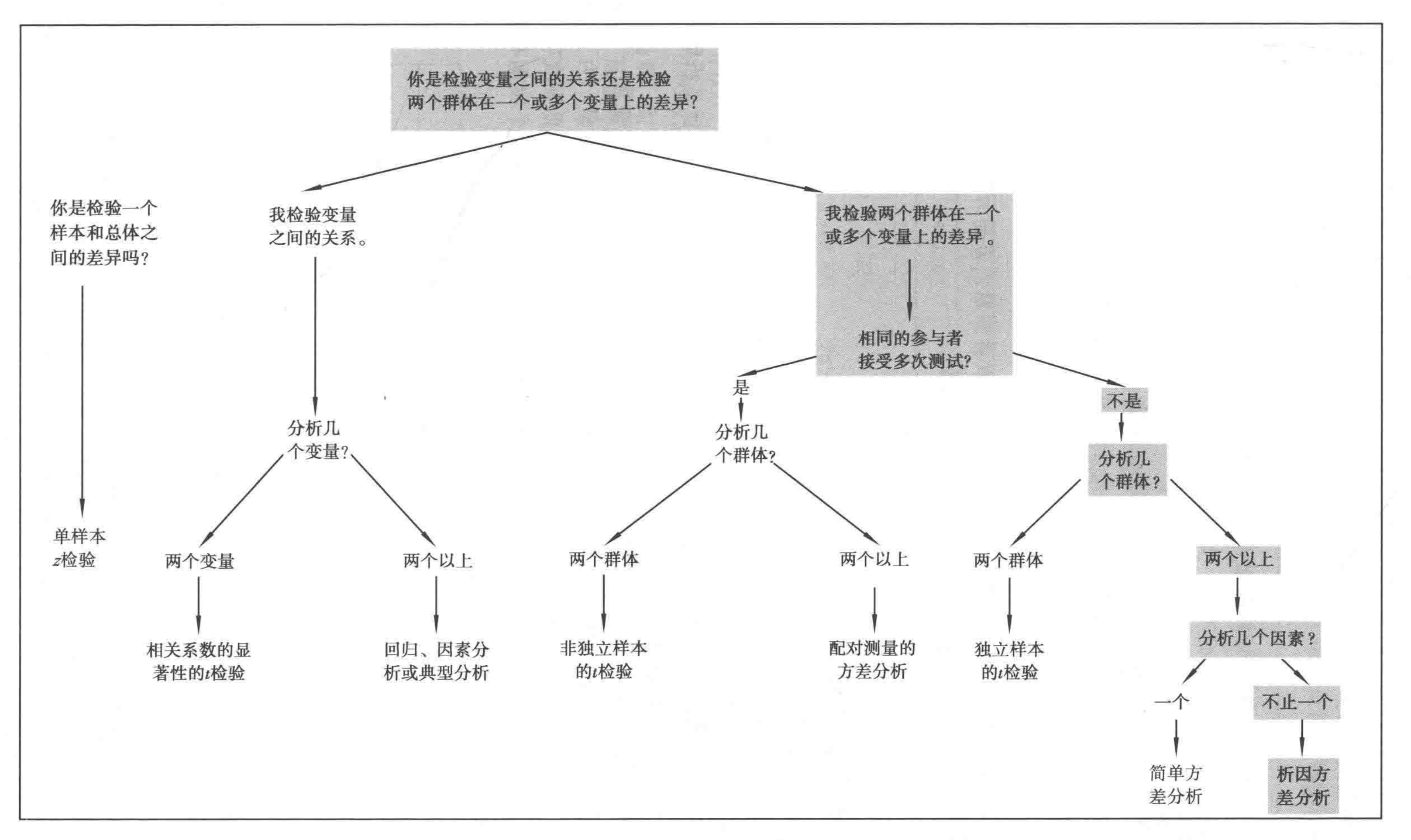


图14.1 确定析因方差分析是正确的统计检验方法

就如在第 13 章一样，我们已经确定方差分析是正确的选择（检验多个群体之间差异或者独立变量的多个层级之间的差异），但是我们要处理不止一个因素，析因方差分析是正确的选择。



1—我们检验不同群体的数值之间的差异。
2—参与者接受不止一次测试。
3—我们检验两个或多个群体。
4—我们处理不止一个因素或独立变量。
5—合适的检验统计是析因方差分析，在交叉条件下复制或重复。



重复方差分析也是指重复测量方差分析或者测量内方差分析，因为测量是对相同参与者（或者其他研究对象）的重复测量。

方差分析的新类型

你已经了解方差分析的一种形式，也就是第 13 章讨论的简单方差分析。简单方差分析只分析一个因素或者处理变量（如群体身份），而且这个因素或者处理变量可分为两个以上的层级或群体。

现在，我们对整个技术进行发展，进而可以同时分析不止一个因素，也就是**析因方差分析**（factorial analysis of variance）。

现在我们来看一个简单的包含两个因素的案例：性别（男和女）和处理变量——处理变量是不同类型的训练项目（高强度和低强度）以及结果——体重减少的数量。为了保持平衡，相同的群体既接受了高强度的也接受了低强度的训练项目（一半先从低强度项目开始，然后是高强度项目；另一半反之）。下面的表格是实验设计。要记住这个案例有重复的方差分析，强度处理变量是重复测量的。

		训练项目	
		高强度	低强度
性别	男性		
	女性		



方差分析设计中包含的一个因素是重复测量而另一个因素不是重复测量，这样的设计有时也叫作混合设计。

接着我们来看什么是主效应和交互效应。现在不进行更多的数据分析，在本章的最后才会相应的增加数据分析，现在只是看和了解。即使在结果中提供了第一类错误的准确概率（我们不使用“$p<0.05$”这样的陈述），我们还是使用 0.05 作为拒绝零假设的标准。

就析因分析来说，你可以提出并回答三个问题。

1.不同的训练项目层级也就是高强度和低强度之间是否有差异？

2.不同性别——男性和女性之间是否有差异？

3.高强度或低强度项目是否对男性或女性有不同的效应？（这就是你将要学习的重要的交互效应。）

问题1和问题2确定主效应是否存在，问题3确定两个因素之间是否有交互效应。

主要方面：析因方差分析中的主效应

你应该记得方差分析的最初目标是检验两个或两个以上群体之间的差异。如果数据分析表明某个因素的不同层级之间存在差异，我们就会说存在**主效应**（main effect）。现在来看一个例子，上面给出的案例中共有四个群体，每个群体10个参与者，总共是40。下面给出可能的分析结果。这个表也叫作**源表**（source table）。

来源	平方和	*df*	均方	*F*	*Sig.*（显著性）
强度	429.025	1	429.025	3.011	0.091
性别	3 222.025	1	3 222.025	22.612	0.000
强度×性别	27.225	1	27.225	0.191	0.665
误差	5 129.700	36	142.492		

现在只关注来源列和*Sig.*列（已经用阴影显示）。我们可以得出的结论是性别有主效应（$p=0.000$），而处理变量没有主效应（$p=0.091$），而且这两个主要因素之间没有交互效应（$p=0.665$）。因此在高强度组或低强度组对减少体重来说并不重要，但是性别是重要的影响因素。而且处理因素和性别之间没有交互效应，所以不同的强度对性别没有差异性影响。如果将对应的均值绘图，就会得到图14.2的形象展示。

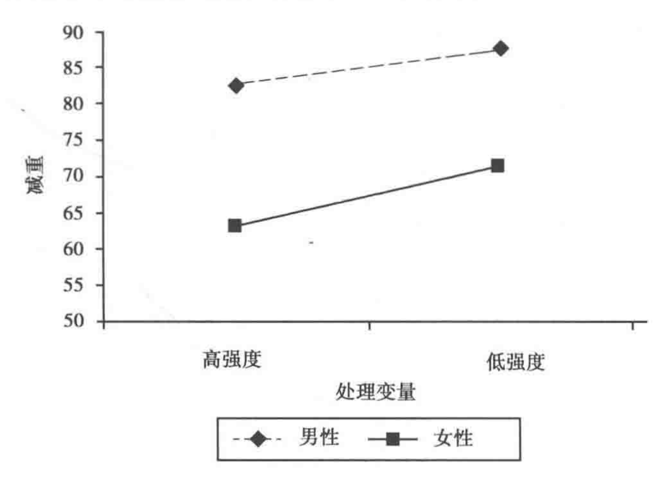


图14.2　处理变量对应男性和女性的均值

在图14.2中你可以看到男性和女性在“减重”轴相差的距离很大（男性的均值是85.25，女性的均值是67.30）。但是对处理变量来说（如果你计算了平均值），你会发现差异很小（高强度项目的均值是73.00，低强度的均值是79.55）。当然现在是进行方差分析，群体内的变化很重要，但是在这个案例中，你可以看到每一个因素（如性别）的群体间的差

异（如男性和女性），以及它们如何在分析结果中反映。

更有趣的方面：交互效应

现在开始讨论交互效应。先看一组新数据的源表，这些数据表明男性和女性受到处理变量的影响，也就是存在**交互效应**（interaction effect）。而且你还可以看到一些非常有趣的分析结果。

来源	平方和	*df*	均方	*F*	*Sig.*
处理变量	265.225	1	265.225	2.444	0.127
性别	207.025	1	207.025	1.908	0.176
处理变量×性别	1 050.625	1	1 050.625	9.683	0.004
误差	3 906.100	36	108.503		
总和	224 321.000	39			

在 0.05 的显著水平下处理变量和性别没有主效应（分别为 $p=0.127$，$p=0.176$），但是存在交互效应（$p=0.004$），这是一个非常有意思的结果。实际上，你是在高强度项目还是低强度项目或者你是男性还是女性都不重要，但是同时处在两种条件下就很重要，如处理变量对于男性和女性的体重减少有不同的影响。

下面的表中给出了四个群体的均值。

	男性	女性
高强度	73.70	79.40
低强度	78.80	64.00

现在给出四个群体的均值的图形展示（图 14.3 所示），都是使用 Excel 的 AVERAGE 函数完成的。

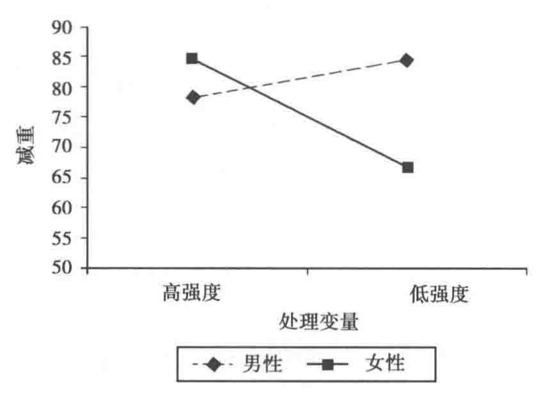


图 14.3 处理变量对男性和女性的均值

如何理解这个结果？解释非常简洁。下面是我们可以做出的解释，就如你可以认识到的，这是之前列出的三个问题的答案。

- 训练类型没有主效应。
- 性别没有主效应。

• 处理变量和性别之间有明显的交互效应，也就是说在高强度项目中女性减少的体重比男性减少的体重多，而在低强度项目中男性减少的体重比女性减少的体重多。

需要记忆的内容



这是特别需要记忆的内容。如果没有掌握更多的知识(也没有持续学习到本章)，你可能会认为你必须做的所有分析都是男性和女性均值之间的 t 检验，或者参与高强度训练项目的群体和参与低强度项目的群体的均值之间的另一种简单的 t 检验——不过你不会有所发现。但是如果考虑到主要因素之间的交互效应，就会发现不同的效应的存在——这是其他的分析没有注意到的结果。实际上你花足够工夫真正搞懂了，就会发现交互效应在任何析因方差分析中都是最有趣的结果。

使用奇妙的分析工具库计算方差分析 F 统计量

现在要面对一种改变。在《爱上统计学：Excel》整本书中，我们提供一些案例展示如何使用老的方式(使用计算器手动计算)或者应用类似 Excel 的统计分析软件包进行特定的统计分析。但对于析因方差分析的介绍，我们只使用 Excel 来说明——这不是由于应用计算器完成析因 ANOVA 更具知识上的挑战性，而是由于计算工作量太大。正是由于这个原因我们不准备手动进行所有的计算，而是直接对重要的值进行计算，并且将更多的时间用在解释方面。

我们使用之前展示的存在显著的交互效应的那组数据。

高强度 男性	高强度 女性	低强度 男性	低强度 女性
76	65	88	65
78	90	76	67
76	65	76	67
76	90	76	87
76	65	56	78
74	90	76	56
74	90	76	54
76	79	98	56
76	70	88	54
55	90	78	56

下面是 F 统计检验量的计算和步骤。这里没有“著名的八个步骤”这个标题，这是全书唯一不用手动计算而只用计算机计算的统计值。因为其分析过程比较复杂，对刚入门的学生来说需要花费的精力太多了。

在我们开始使用分析工具库的析因方差分析或者二元方差分析工具之前，首先我们陈述零假设和研究假设。实际上这里出现的零假设有三个(公式 14.1a，14.1b，14.1c)，表明两个因素的均值之间没有差异，没有交互效应。现在就具体描述。

首先，对处理变量来说，

$$H_0: \mu_{high} = \mu_{low} \tag{14.1a}$$

对性别变量来说，

$$H_0: \mu_{male} = \mu_{female} \quad (14.1b)$$

对处理变量和性别之间的交互效应来说，

$$H_0: \mu_{high \cdot male} = \mu_{high \cdot female} = \mu_{low \cdot male} = \mu_{low \cdot female} \quad (14.1c)$$

公式 14.2a，14.2b，14.2c 所示为研究假设，表明群体的均值之间有差异，而且存在交互效应。具体如下：

首先，对处理变量来说，

$$H_1: \bar{\mu}_{high} \neq \bar{\mu}_{low} \quad (14.2a)$$

对性别变量来说，

$$H_1: \bar{\mu}_{male} \neq \bar{\mu}_{female} \quad (14.2b)$$

对处理变量和性别之间的交互效应来说，

$$H_1: \bar{\mu}_{high \cdot male} \neq \bar{\mu}_{high \cdot female} \neq \bar{\mu}_{low \cdot male} \neq \bar{\mu}_{low \cdot female} \quad (14.2c)$$

我们使用方差分析：可重复双因素方差分析选项，然后按下面的步骤进行。上面所示的数据可以在附录 C 找到，就是第 14 章数据集 1，也在图 14.4 中显示。如果你想实际操作使用分析工具库，那么确保数据文件已经打开，或者输入数据并保存文件。

	A	B	C
1		High	Low
2	Male	76	88
3		78	76
4		76	76
5		76	76
6		76	56
7		74	76
8		74	76
9		76	98
10		76	88
11		55	78
12	Female	65	65
13		90	67
14		65	67
15		90	87
16		65	78
17		90	56
18		90	54
19		79	56
20		70	54
21		90	56

图 14.4　用于方差分析样例数据——第 14 章数据集 1

1.点击数据→数据分析，你可以看到数据分析对话框。简单地回顾一下如何使用分析工具库？第 1 章 1b 部分。

2.点击方差分析：可重复双因素方差分析，接着点击确定，你会看到图 14.5 所示的对话框。

3.在输入区域对话框，输入你想让 Excel 在计算方差时使用的数据，如图 14.4 所示，我们想分析的数据的行和列的范围是从 A1 到 C21。

4.在每一样本的行输入数字。这个数字与观察对象的数量一致，是 10。

5.输入你确定的 F 值检验的显著水平。在这个案例中，我们使用的显著水平是 $p<0.05$。

6.在对话框的输出选项部分点击输出区域选项，输入你想让 Excel 将分析结果返回的位置。在这个案例中，我们选择单元格 E1。输入完成的对话框见图 14.6。

7.点击确定。大家就可以看到图 14.7 所示的输出结果。现在真的应该感到高兴。（我们做了一些编辑，如四舍五入，图 14.7 才呈现出更容易理解的形式。）

Excel 没有给出方差来源的标签，而是使用通用的词汇。因此在这个案例中，“Sample”表示性别，“Columns”表示项目强度，“Interaction”表示项目强度和性别之间的交互效应。

Anova: Two-Factor With Replication

Input

Input Range:

Rows per sample:

Alpha: 0.05

OK

Cancel

Help

Output options

Output Range:

New Worksheet Ply:

New Workbook

图 14.5　有重复的两因素方差分析对话框

Anova: Two-Factor With Replication

Input

Input Range: $A$1:$C$21

Rows per sample: 10

Alpha: 0.05

OK

Cancel

Help

Output options

Output Range: $E$1

New Worksheet Ply:

New Workbook

图 14.6　完成的方差分析对话框

	A	B	C	D	E	F	G	H	I	J	K
1		High Impact	Low Impact		Anova: Two-Factor With Replication						
2	Male	76	88								
3		78	76		SUMMARY	High Impact	Low Impact	Total			
4		76	76		*Male*						
5		76	76		Count	10	10	20			
6		76	56		Sum	737	788	1525			
7		74	76		Average	73.7	78.8	76.25			
8		74	76		Variance	44.46	121.96	85.67			
9		76	98								
10		76	88		*Female*						
11		55	78		Count	10	10	20			
12	Female	65	65		Sum	794	640	1434			
13		90	67		Average	79.40	64	71.70			
14		65	67		Variance	141.38	126.22	189.17			
15		90	87								
16		65	78		*Total*						
17		90	56		Count	20	20				
18		90	54		Sum	1531	1428				
19		79	56		Average	76.55	71.40				
20		70	54		Variance	96.58	175.2				
21		90	56								
22											
23					ANOVA						
24					*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
25					Sample	207.03	1	207.03	1.91	0.18	4.11
26					Columns	265.22	1	265.22	2.44	0.13	4.11
27					Interaction	1050.63	1	1050.63	9.68	0.00	4.11
28					Within	3906.10	36	108.50			
29											
30					Total	5428.98	39				
31					*F crit*						
32					2.17						
33					4.38						

图 14.7　使用 Excel 分析工具库完成的二元方差分析

Tech Talk　你在这里学习的分析是单变量方差分析。这项分析只处理一个被解释变量或者结果变量——在我们的这个案例中就是体重减少。如果我们的研究问题中不止一个被解释变量(如对饮食的态度),那么就是多变量方差分析(也翻译为多元方差分析),在检验群体差异的同时要控制被解释变量之间的关系。第 18 章会讨论更多的内容。

小　结

我们已经完成了群体均值之间差异的检验，接下来要研究相关关系，或者说两个变量之间的关系的显著性。

练　习

1.说明何时使用析因方差分析而不是简单方差分析检验两个或更多群体之间差异的显著性。

2.设计一项可以进行析因方差分析的2×3实验。

3.使用第14章数据集2完成分析并解释结果。这是一个2×3的实验设计：严重程度（severity）有两个层级，一个层级是严重（severe），一个层级是中等（mild）；处理变量有三个类型，一是药物1，二是药物2，三是安慰剂。这是一个有重复方差分析，因为每个参与者要接受所有的治疗，由列数据表示；严重程度是用行数据表示（或者是Excel的数据分析结果中被叫作"Sample"的）。

4.第14章数据集3中的数据表示一项试验的结果，这项实验设计是：营养教育有三个层级、饮食干预有两个层级。要检验其效果，效果使用100分的量表，100分表示最高的健康饮食水平。回答下面的问题。

a.饮食干预是否有效果？

b.营养教育是否很重要？

c.两个主效应之间是否有交互效应？

d.你的结论是什么？（用一句话来概括。）

是近亲还是好朋友——使用相关系数检验关系 15

本章你会学到什么 ☺☺☺☺ ☒☒

- 如何检验相关系数的显著性
- 相关系数的解释
- 如何使用 PEARSON 函数
- 统计显著性和结论有意义的重要区别

相关系数检验的介绍

丹尼尔·舍克(Daniel Shek)在他研究婚姻质量与父母-子女关系的文章中告诉我们至少存在两种可能性。首先,失败的婚姻可能促进父母-子女关系。这是由于父母对婚姻不满意,就可能以与子女的关系替代夫妻关系来实现情感的满足。另外,依据进一步的假设,失败的婚姻也可能破坏父母-子女关系。这是由于失败的婚姻可能增加了抚养子女的困难。

舍克研究了 378 对中国夫妇 2 年期内的婚姻质量与父母-子女关系之间的关系。他发现婚姻质量水平越高,父母-子女关系水平也越高;截面(现时)数据和纵向(过去一段时期)数据都支持这项发现。他也发现父母-子女关系的强度对父亲和母亲是一样的。这是一个明显的例证,可以说明如何使用相关系数给出我们需要的关于一个变量和另一个变量相关的信息。舍克全面计算了时期 1、时期 2 与父亲和母亲之间不同的相关关系,所有的计算的目的都相同:确定变量之间是否显著相关。要记住,这不是说变量关系存在任何因果性,只是说一个变量和另一个变量相关。

如果想了解更多,请查阅舍克(Shek,D.T.L.)1998 年发表在《家庭问题杂志》(*Journal of Family Issues*)第 19 期 687~704 页的文章“婚姻质量和父母-子女关系之间的关系(Linkage between marital quality and parent-child relationship: A longitudinal study in the Chinese culture.)。”

智慧和知识之路

下面介绍如何使用流程图选择合适的统计检验来检验相关系数。沿着图 15.1 中指示的步骤就可以。

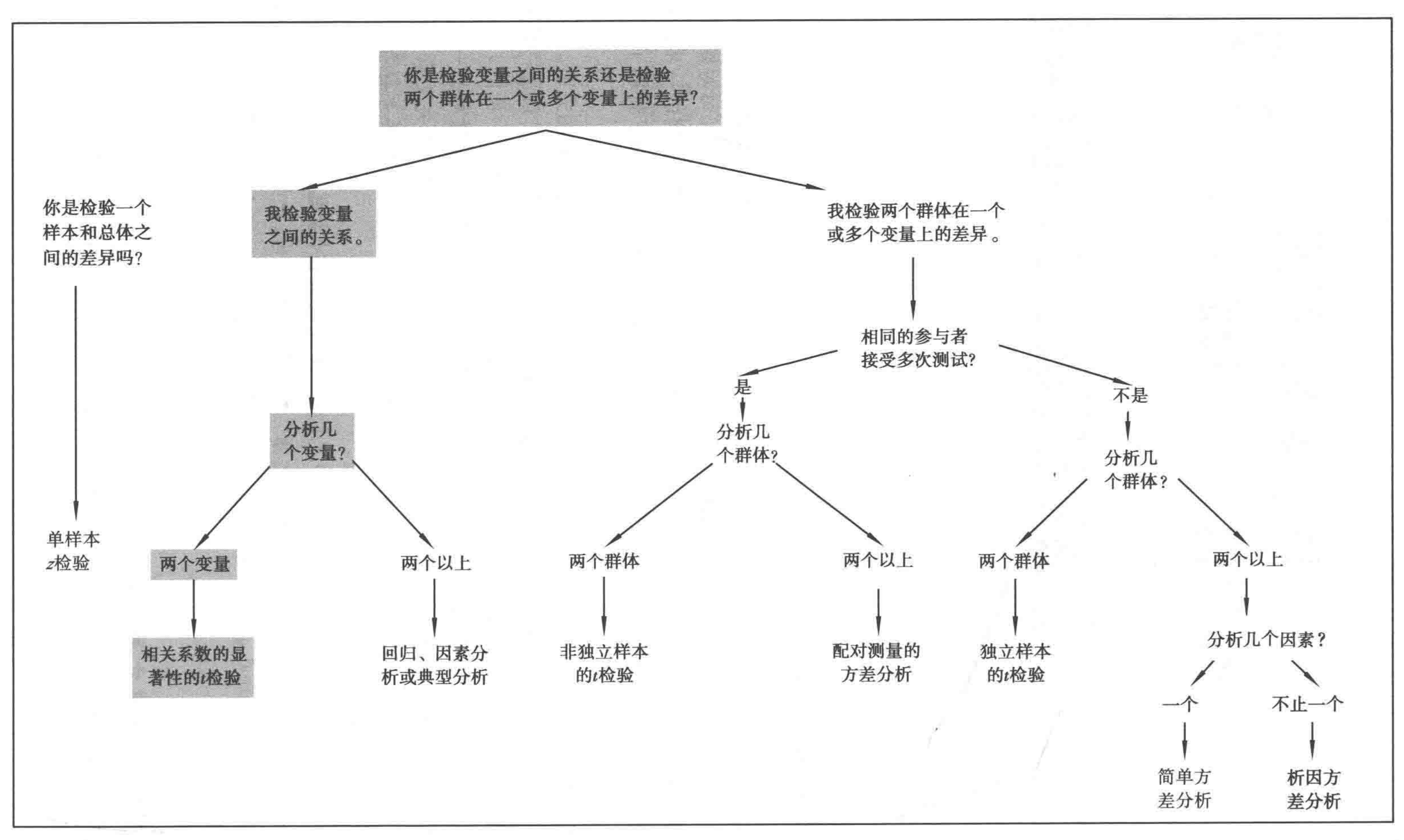


图 15.1　确定相关系数 t 检验是合适的统计检验方法



1—我们检验变量之间的关系而不是群体之间的差异。
2—只检验两个变量之间的关系。
3—所要应用的合适的检验统计是相关系数的 t 检验。

计算检验统计量

你会喜欢看到下面的资料:相关系数本身可以作为自己的检验统计量。你不需要计算任何检验统计量,而且检验显著性也非常容易,这使事情变得相当简单。我们现在利用下面的一组数据举例说明两个变量,也就是婚姻质量和父母-子女关系水平之间关系的检验,

婚姻质量	父母-子女关系水平	婚姻质量	父母-子女关系水平
76	43	59	38
81	33	87	21
78	23	77	27
76	34	79	43
76	31	85	46
78	51	68	41
76	56	76	41
78	43	77	48
98	44	98	56
88	45	99	55
76	32	98	45
66	33	87	68
44	28	67	54
67	39	78	33
65	31		

可以应用第 5 章的公式 5.1 计算皮尔森相关系数。具体计算时你会得到的结果是 $r=0.393$。现在我们按步骤检验这个值的显著性,并且确定这个值的含义。

更多 Excel

你已经知道如何使用 CORREL(相关)函数,以及分析工具库中的相关工具,这两种工具都很容易计算相关系数。你也应该了解 Excel 提供的另一种相关工具,这就是 PEARSON 函数。这个函数和 CORREL 函数、分析工具库中的相关工具一样计算相关系数。就你们的学习目的来说,这三种方法计算相同的值,而且是可以互换的。当你们选修了更高级的统计课程,你可能会想了解 CORREL 函数和 PEARSON 函数之间有什么不同(答案是是否具有显著性检验)。

下面是著名的八个步骤和检验统计量的计算。

1.零假设和研究假设的表述。零假设是婚姻质量和父母-子女关系之间没有关系。研究假设是双侧的无方向的假设,因为研究假设只是陈述两个变量之间有关系,而方向并不重要。记住相关可能是正向的或负向的,但相关系数最重要的特征是它的绝对值大小而

不是符号（正或者负）。

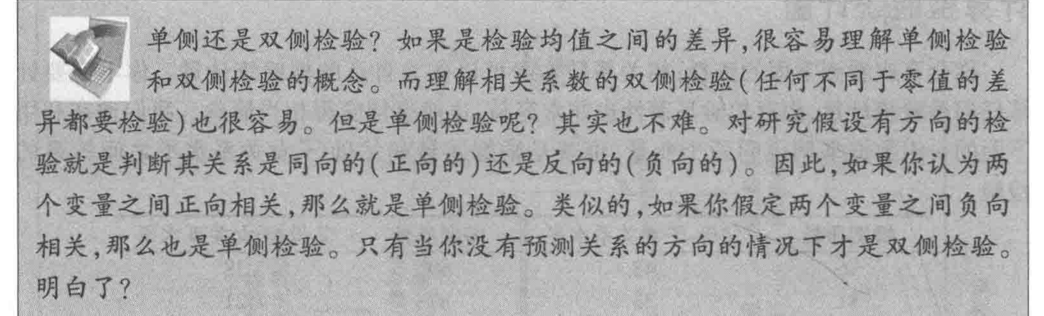
零假设如公式 15.1 所示：

$$H_0: r_{xy}=0 \tag{15.1}$$

希腊字母 r 或 rho 表示相关系数的总体估计。

研究假设（公式 15.2 所示）表示两组值之间有关系，而且这个关系值不等于 0。

$$H_1: r_{xy}\neq 0 \tag{15.2}$$



单侧还是双侧检验？如果是检验均值之间的差异，很容易理解单侧检验和双侧检验的概念。而理解相关系数的双侧检验（任何不同于零值的差异都要检验）也很容易。但是单侧检验呢？其实也不难。对研究假设有方向的检验就是判断其关系是同向的（正向的）还是反向的（负向的）。因此，如果你认为两个变量之间正向相关，那么就是单侧检验。类似的，如果你假定两个变量之间负向相关，那么也是单侧检验。只有当你没有预测关系的方向的情况下才是双侧检验。明白了？

2.设置零假设的风险水平（显著性水平，或第一类错误）。风险是 0.05。

3.和 4. 选择合适的检验统计量。使用图 15.1 所示的流程图，我们确定用于相关系数的合适的检验。在这个案例中，我们不需要计算检验统计量，因为简单的 r 值（$r_{xy}=0.393$）就我们的检验目的来说就是我们的检验统计量。

5.使用特定统计量的合适的临界值表来确定拒绝零假设需要的值。附录 B 的表 B4 列出了相关系数的临界值。

首先我们要确定自由度（df），自由度近似样本规模。就现在的检验统计量来说，自由度是 $n-2$，或者 $30-2=28$，其中 n 等于用于计算相关系数的配对数量。这里的自由度只适合于这项统计检验，但不是其他统计检验所需要的。

使用自由度数值（28）、你愿意承担的风险水平（0.05）和双侧检验（因为研究假设没有方向），临界值是 0.381（使用 $df=25$，25 是最接近的数字）。因此对显著水平为 0.05、自由度为 28 的双侧检验来说，我们拒绝零假设需要的值是 0.381。



好了，我们有点儿投机取巧。实际上你可以计算 t 值（类似于不同均值之间差异的检验）检验相关系数的显著性。计算公式并不比之前几章学到的公式难，但是你不会在本章看到。重点是一些聪明的统计学家已经计算了在不同显著水平下（0.01 和 0.05）不同样本规模的单侧检验和双侧检验临界 r 值，就如表 B4 所示。如果你在读专业期刊时看到过使用 t 值检验相关关系，现在你知道这是为什么了吧。

6.比较实际值和临界值。实际值是 0.393，拒绝零假设也就是两个变量不相关的临界值是 0.381。

7.和 8.做出决定。如果实际值（或者检验统计量的值）大于临界值（表中列出的值）就不能接受零假设。如果实际值没有超过临界值，零假设就是最有力的解释。

在这个案例中，实际值（0.393）超过临界值（0.381）——这个值足够大，我们可以说两个变量（婚姻质量和父母-子女关系）之间确实存在除随机因素之外的某些因素引起的某

种关系。

如何解释 $r_{28}=0.393, p<0.05$

- r 表示我们使用的检验统计量。
- 28 是自由度数值。
- 0.393 是实际值，是使用第 5 章给出的公式计算所得的值。
- $p<0.05$（p 值是最重要的）表示对零假设的任何检验来说，两个变量之间的关系缘于随机因素的可能性小于 5%。因为我们以 0.05 作为研究假设比零假设更有解释力的标准，我们的结论就是两个变量之间的关系是显著的。这意味着随着婚姻质量水平的提高父母-子女关系的水平也提高。相应的，随着婚姻质量水平的降低父母-子女关系的水平也降低。

回忆第 6 章的内容，我们使用相关系数评价不同类型的测量工具的信度和效度。相关系数也会用于更复杂的统计技术，一般是进行分类数据的融合（如因子分析）。因此，就算你还没有掌握相关系数，（即使相关系数有时被过度解释，还记得冰激凌和犯罪之间的关系？）也一定要知道相关系数非常有用也经常要用到。

（再次说明）因果和相关

你可能认为你已经对因果和相关有了足够的了解，但是这一点太重要了，因此我们不得不再申明一次。仅仅因为两个变量相关（类似上面给出的案例），并不表示一个变量变化就引起另一个变量的变化。换句话说，婚姻质量高并不能保证父母-子女关系水平也高。这两个变量可能因为共享一些使得一个人成为好丈夫、好妻子或好父母的特质（耐心、领悟力、牺牲的意愿）而相关，但是也有相当的可能看到很多好丈夫或好妻子与子女的关系相当糟糕。

是否还记得第 5 章的犯罪和冰激凌的案例？在本章也一样。仅仅因为变量相关并分享一些共同的特征，并不意味着这两个变量之间存在因果关系。

再强调一遍显著性和意义

在第 5 章我们对使用决定系数理解相关系数的意义进行了评述。你可能记得相关系数的平方可用于确定一个变量的方差可以解释另一个变量方差的程度。在第 9 章我们也讨论了显著性与意义的问题。

但是我们需要再一次讨论这个主题。即使相关系数是显著的（就如本章的案例），并不意味着可以解释的方差的程度是有意义的。例如，在这个案例中，简单皮尔森相关值 0.393的决定系数等于 0.154，这表明可以解释方差的 15.4%，而方差的 84.6% 不能解释。这就留下了很大的质疑空间，对吧？

因此，即使我们知道婚姻质量和父母-子女关系之间有正向的关系，这两个变量可能“走”在一起，但是这么小的相关系数 0.393 表明在这两个变量的关系中还有其他很重要

的因素在发挥作用。因此，我得告诉你统计学中的一句谚语，"你看到的并不总是你得到的"。

小 结

相关系数是指出关系的方向，并帮助我们更好地理解两个结果变量共享内容的重要工具。要记住相关只适用于相互关系而不是因果关系。

练 习

1.参照下面给出的信息，使用附录B的表B4确定相关系数是否显著，以及如何解释结果。

a.20个妇女的速度和力量之间的相关系数是0.567。在0.05的显著水平下使用单侧检验来检验结果。

b.数学考试中回答正确的问题数量和完成考试的时间之间相关系数是-0.45。在0.05的显著水平下检验80个孩子的相关系数是否显著。选择单侧检验还是双侧检验？并给出解释。

c.50个青少年的朋友数量和平均成绩(GPA)之间的相关系数为0.37。在0.05的显著水平下双侧检验是否显著？

2.使用第15章数据集2回答下面的问题。手动计算或者使用工具库。

a.计算动机和GPA之间的相关系数。

b.使用双侧检验在0.05的显著水平下检验相关系数是否显著。

c.正确或错误？动机水平越高就越愿意学习。你的选择是什么并解释理由。

3.讨论并举例说明(不要再用冰激凌和犯罪的例子)：两个变量相关并不意味着一个变量的变化会引起另一个变量的变化。

4.使用第15章数据集2，使用CORREL函数或者PEARSON函数计家庭图书数量和一般教育水平(或者完成的教育年限，12年就等于高中毕业)之间的相关系数。基于结果，如果你想让一个人能上研究所并获得博士学位，你就得将成千上万的图书搁在他们的书架上，对吗？

预测谁将赢得超级杯——使用线性回归 16

本章你会学到什么

- 如何进行估计,估计如何应用在社会科学和行为科学中
- 以一个变量估计另一个变量时如何以及为什么进行线性回归
- 如何判断估计的准确性
- 如何使用 INTERCEPT 和 SLOPE 函数
- 如何使用分析工具库中的回归工具
- 多元回归及其应用

什么是估计

现在简短地介绍一下。你不仅可以计算两个变量相关的程度(就如在第 5 章一样计算相关系数),也能够以相关系数为基础,通过一个变量的值计算另一个变量的值。这是应用相关的非常特别的案例,而且是社会和行为科学研究者的强力工具。

估计的基本含义就是使用已经收集的数据集(如变量 X、Y 的数据),计算变量如何相关,接着使用相关系数以及 X 的信息来估计 Y。听起来有些难? 实际上不难,特别是当你通过下面的例子对估计有了了解之后。

例如,研究者收集了 400 个州立大学的学生高中时期的 GPA 成绩和大学第一年的 GPA。他计算了这两个变量之间的相关系数。接着他使用你将在本章学到的统计技术,利用新的 400 个学生样本的高中 GPA(已经从之前的学生数据集了解高中 GPA 与大学第一年 GPA 有关系)估计大学第一年的 GPA。很好用,对吧?

这儿还有另一个案例。一些教师对留级发挥的作用感兴趣。也就是一些孩子留在幼儿园(没有进入一年级)是否比进入一年级更好? 这些教师知道留级和一年级的成绩的相关系数,他们可以将相关系数应用到新的学生样本并依据幼儿园的成绩估计一年级的成绩。这样行得通吗? 行得通。收集过去事件(如两个变量之间现存的关系)的数据,然后在只知道一个变量的情况下应用到未来的事件中。这要比你想象的容易。



相关系数的绝对值越大，依据相关以一个变量估计另一个变量的准确性越高，因为这两个变量共享的部分越多，依据对第一个变量的了解就可以更多地了解第二个变量。而且你可能已经得出结论，也就是如果相关是完全相关（+1.0 或-1.0），那么估计也是完全估计。如果 $r_{xy}=-1.0$ 或 1.0，而你知道 X 值，那么你就可以知道 Y 值。当然，如果 $r_{xy}=-1.0$ 或 1.0，而你知道 Y 值，那么你也就可以知道 X 值。这两种估计形式都可以发挥很好的效用。

本章的内容是详细了解线性回归过程，学会通过 X 值估计 Y 值。我们从估计的一般逻辑的讨论开始，接着介绍一些简单的绘制线图的技能，最后以具体的案例讨论估计过程。

估计的逻辑

在开始实际的计算并展示如何使用相关进行估计之前，我们先讨论为什么以及如何估计。接着我们继续使用以高中 GPA 估计大学 GPA 的案例。

估计是以过去的结果推测未来结果的活动。当我们想通过一个变量估计另一个变量，我们首先需要计算两个变量之间的相关系数。表 16.1 给出这个案例中使用的数据。图 16.1 是将要计算的两个变量的散点图。

表 16.1　高中 GPA 和大学第一年 GPA 数据

高中 GPA	大学第一年 GPA
3.50	3.30
2.50	2.20
4.00	3.50
3.80	2.70
2.80	3.50
1.90	2.00
3.20	3.10
3.70	3.40
2.70	1.90
3.30	3.70

要以高中 GPA 估计大学 GPA，我们必须建立**回归方程**（regression equation），并使用这个等式建立**回归线**（regression line）。回归线反映我们以变量 X 的值（高中 GPA）估计变量 Y 值（大学 GPA）的最好猜测。你会很快学会如何绘制图 16.2 所示的回归线。回归线表示什么？

首先这是变量 Y 对变量 X 的回归。换句话说，Y（大学 GPA）将依据 X（高中 GPA）被估计。回归线也称为**最优拟合线**（line of best fit）。回归线最好地拟合了数据，因为这条线将每个数据点与回归线的距离最小化。例如，你考虑所有的数据点并尽量寻找可以同时拟合这些数据点的直线，你会看到图 16.2 中所示的直线就是你要用到的直线。

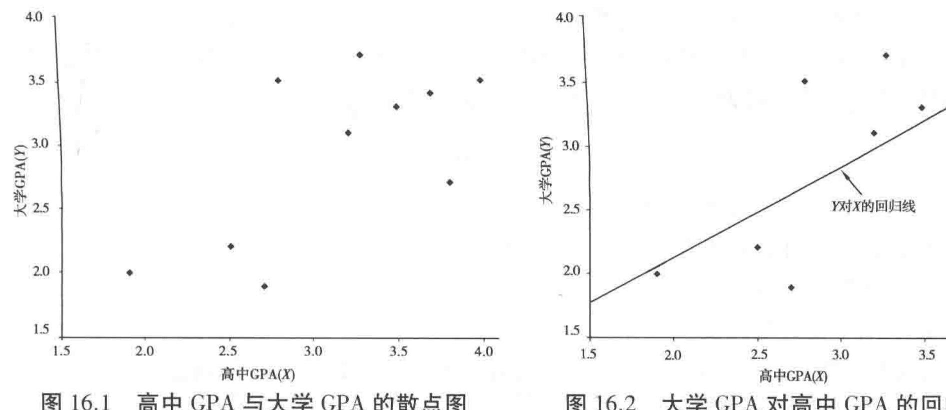


图 16.1 高中 GPA 与大学 GPA 的散点图

图 16.2 大学 GPA 对高中 GPA 的回归线

其次,这条直线可以用于进行最好的猜测(已知高中 GPA 的情况下对大学 GPA 进行估计)。例如,如果高中 GPA 是 3.0,那么大学 GPA 就大概(记住,这只是看图示大概估计)是 2.8。看图 16.3 你就会了解我们是怎么估计的。我们先在 x 轴定位估计值(3.0),然后绘制从 x 轴到回归线的垂直线,接着绘制水平线到 y 轴并估计对应值的大小。

第三,每一个具体数据点和回归线的距离就是**估计误差**(error in prediction)——是两个变量之间相关的直接反映。例如,你看到的数据点(3.3,3.7)(图 16.4 中所示),在回归线的上方。这个数据点和回归线的距离就是估计误差,就如图 16.4 中所标注的一样。那么如果是完全估计,所有估计的数据点会落在什么位置?刚好就在回归线或估计线上。

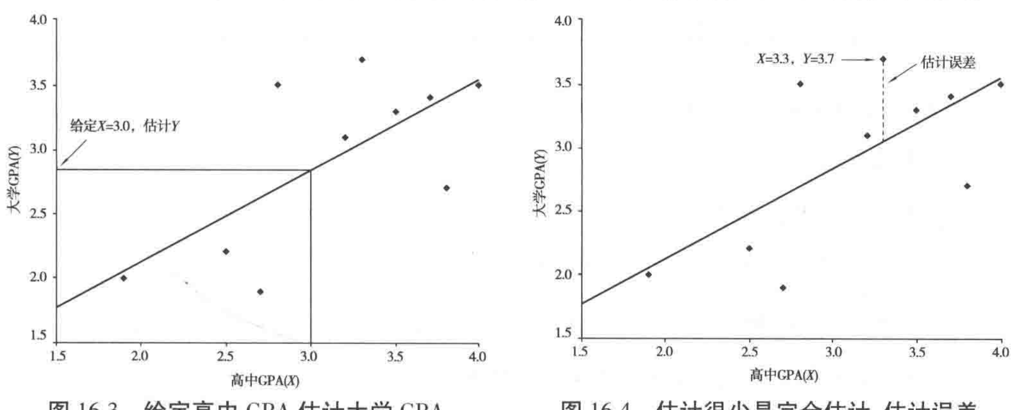


图 16.3 给定高中 GPA 估计大学 GPA

图 16.4 估计很少是完全估计:估计误差

第四,如果是完全相关,所有的数据点将沿着 45°角成为一条直线,而且回归线通过每一个数据点(就如上面的第三点所讲)。

给定回归线之后,我们可以使用回归线估计任何未来的值。这正是我们要做的——建立回归线然后进行估计。

绘制拟合数据的最优直线

理解估计的最简单的方式就是依据一个变量值[我们称为 X——自变量(independent variable)或估计变量(predictor)]确定另一个变量的值[我们称为 Y——因变量(dependent variable)或标准变量(criterion)]。

我们发现利用 X 估计 Y 的很好的方式就是建立本章之前提到的回归线。回归线依据

我们已经收集的数据建立。接着利用等式以 X 或估计变量的新的数值进行估计。

公式 16.1 所示是回归线的一般公式，这个公式在高中或者大学的数学课程中都可能用到，对你来说应该很熟悉。这个公式与其他任何直线公式一样。

$$Y' = bX + a \tag{16.1}$$

其中：Y' 表示已知 X 值的 Y 的估计值，

b 表示直线的斜率或者方向，

a 表示直线与 y 轴相交的点，

X 表示用于估计的数值。

现在我们使用之前表 16.1 中列出的相同数据进行一些必要的计算。

	X	Y	X^2	Y^2	XY
	3.5	3.3	12.25	10.89	11.55
	2.5	2.2	6.25	4.84	5.50
	4.0	3.5	16.00	12.25	14.00
	3.8	2.7	14.44	7.29	10.26
	2.8	3.5	7.84	12.25	9.80
	1.9	2.0	3.61	4.00	3.80
	3.2	3.1	10.24	9.61	9.92
	3.7	3.4	13.69	11.56	12.58
	2.7	1.9	7.29	3.61	5.13
	3.3	3.7	10.89	13.69	12.21
总计	31.4	29.3	102.50	89.99	94.75

这些是表格中最下一行各总计的含义：

$\sum X$ 或所有 X 值的总和是 31.4。

$\sum Y$ 或所有 Y 值的总和是 29.3。

$\sum X^2$ 或每个 X 值的平方和是 102.50。

$\sum Y^2$ 或每个 Y 值的平方和是 89.99。

$\sum XY$ 或所有 X 值和 Y 值乘积的和是 94.75。

公式 16.2 用于计算回归线的斜率（直线公式中的 b）：

$$b = \frac{\sum XY - (\sum X \sum Y/n)}{\sum X^2 - [(\sum X)^2/n]} \tag{16.2}$$

在公式 16.3 中你可以看到直线斜率 b 的计算值。

$$b = \frac{94.75 - [(31.4 \times 29.3)/10]}{102.5 - [(31.4)^2/10]} \tag{16.3}$$

$$b = \frac{2.74}{3.904} = 0.702$$

公式 16.4 用于计算直线与 y 轴相交的点（直线公式中的 a）：

$$a=\frac{\sum Y-b\sum X}{n} \tag{16.4}$$

在公式 16.5 中你可以看到直线截距 a 的计算值。

$$a=\frac{23.3-0.702\times31.4}{10} \tag{16.5}$$

$$a=\frac{7.26}{10}=0.726$$

现在回到直线的等式（$Y'=bX+a$）并代入具体的 a 值和 b 值，最后的回归直线就表示为：

$$Y' = 0.702X + 0.726 \tag{16.6}$$

为什么是 Y' 而不只是 Y？记住，我们用 X 估计 Y，Y'（读作 Y 撇）是估计值而不是实际的 Y 值。现在我们已经建立了等式，接下来我们能做什么？估计 Y，还有其他吗？

例如，我们可以说高中 GPA 等于 2.8（或 $X=2.8$），如果我们将 2.8 这个值代入等式，就会得到下面的公式：

$$Y' = 0.702(2.8) + 0.726 = 2.69 \tag{16.7}$$

因此 2.69 就是 X 等于 2.8 时 Y 的估计值（或 Y'）。因此对任何的 X 值，我们可以很容易地快速计算 Y 的估计值。

更多 Excel

我们可以使用这个公式并使用已知数值来计算想估计的数值。这是我们讨论的大部分内容。但是你也可以绘制回归线来展示我们估计的数值实际上拟合用于估计的数据的程度。如图 16.2 是我们之前所用的高中-大学 GPA 数据的散点图和一条趋势线（回归直线的另一个名称）。我们如何得到这条直线？我们使用第 5 章学习的绘图技术建立散点图，接着我们在布局选项中选择趋势线，并且选择线性趋势。很简单，很快就能完成。

你看到的趋势线是正向的（也就是这条线有正向的斜率），而且相关系数是 0.6835，非常正向。你可以看到数据点并不刚好都在这条线上，但大都非常接近，这表明相应的误差比较小。

Tech Talk

并不是所有的最优拟合数据点的回归线都是直线。回归线可能是曲线，就如我们在第 5 章讨论的，变量间的关系是曲线关系。例如焦虑和成绩之间的关系就是曲线关系，也就是当人们完全不焦虑或者非常焦虑时，他们的成绩都不好。只有他们的焦虑适度，成绩才能最大化。这两个变量之间的关系就是曲线关系，所以在以 X 值估计 Y 值时要考虑。如果所用的数据是曲线关系就增加了解释的复杂性，需要使用不同的分析工具，相关的内容最好留待后面学习。

更多 Excel

Excel 中有大量的函数是分析某种线性回归，例如 TREND、LINEST、FREQUENCY、STEYX 以及 FORECAST 函数（你可以按 F1 功能键通过 Excel 帮助功能了解这些函数的功能以及如何使用）。我们打算只讨论能反映我们本章讨论内容的两个函数。第一个是 SLOPE 函数，这个函数计算直线的斜率，或者公式 $Y' = bX + a$ 中的 b；第二个是 INTERCEPT 函数，这个函数计算相同等式中的 a 或者直线与 y 轴相交的点。

使用 Excel 的 SLOPE 函数

SLOPE 函数计算直线的斜率，或者我们之前讨论的线性回归等式中的 b。按照下面的步骤计算斜率。

1.我们使用图 16.5 所示的数据，代表的是两个数据集。

	A	B
1	High School GPA	First Year College GPA
2	3.5	3.3
3	2.5	2.2
4	4.0	3.5
5	3.8	2.7
6	2.8	3.5
7	1.9	2.0
8	3.2	3.1
9	3.7	3.4
10	2.7	1.9
11	3.3	3.7

图 16.5 用于回归的样本数据

2.选择你想输入 SLOPE 函数的单元格。在这个案例中，我们准备将返回的值放入单元格 B13。我们也给单元格 A13 定义标签，这样显得清晰，也容易理解。

3.点击［公式］→［其他函数］→［统计］→［SLOPE］，你会看到图 16.6 所示函数参数对话框。

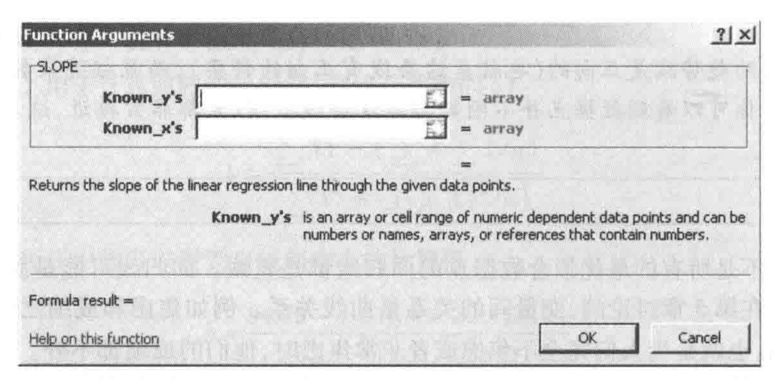


图 16.6 SLOPE 函数参数对话框

4.在 Known_y's 输入单元格范围，这个变量也就是大学的 GPA 成绩或者变量 Y。变量的范围是 B2 到 B11。如图 16.5 所示。

5.在 Known_x's 输入单元格范围，这个变量就是高中 GPA 成绩或者变量 X。变量的范围是 A2 到 A11。图 16.7 所示是完成的对话框形式。

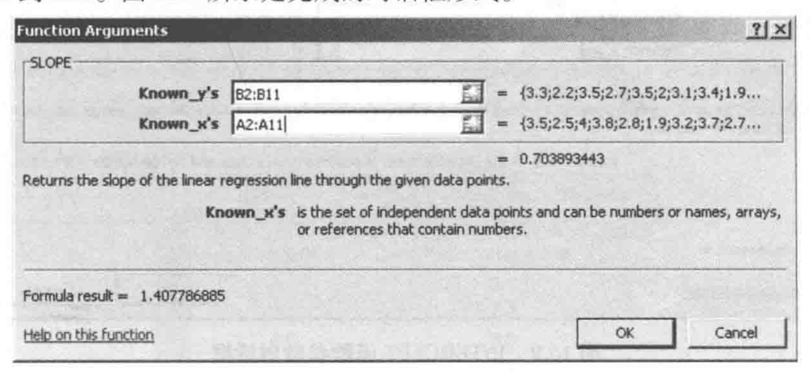


图 16.7 完成的 SLOPE 函数对话框

6.点击确定。你就会在单元格 B13 看到最优拟合回归线的斜率，如图 16.8 所示。你可以看一下在公式栏出现的完整的函数表达式。

B13 =SLOPE(B2:B11,A2:A11)

	A	B
1	High School GPA	First Year College GPA
2	3.5	3.3
3	2.5	2.2
4	4.0	3.5
5	3.8	2.7
6	2.8	3.5
7	1.9	2.0
8	3.2	3.1
9	3.7	3.4
10	2.7	1.9
11	3.3	3.7
12		
13	Slope (b)	0.703893443

图 16.8 完成 SLOPE 函数

使用 Excel 的 INTERCEPT 函数

INTERCEPT 函数计算回归线与 y 轴交叉的位置，也就是我们之前讨论的回归等式中的 a。

按照下面的步骤应用 INTERCEPT 函数。



1.我们使用说明 SLOPE 函数时相同的数据。(见图 16.5)

2.选择你想输入 INTERCEPT 函数的单元格。在这个例子中，我们准备将返回的值放在单元格 B14。我们也在单元格 A14 设置标签，这样显得清晰，也容易理解。

3.点击[公式]→[其他函数]→[统计]→[INTERCEPT]，你会看到图 16.9 所示函数参数对话框。

4.在 Known_y's 输入单元格范围，这个变量也就是大学的 GPA 成绩或者变量 Y。变量的范围是 B2 到 B11。

Function Arguments
INTERCEPT
Known_y's = array
Known_x's = array
=
Calculates the point at which a line will intersect the y-axis by using a best-fit regression line plotted through the known x-values and y-values.
Known_y's is the dependent set of observations or data and can be numbers or names, arrays, or references that contain numbers.
Formula result =
Help on this function OK Cancel

图 16.9 INTERCEPT 函数参数对话框

5.在 Known_x' s 输入单元格范围,这个变量就是高中 GPA 成绩或者变量 X。变量的范围是 A2 到 A11。图 16.10 所示是完成的对话框形式。

6.点击确定。你就会在单元格 B13 看到最优拟合回归线的截距,如图 16.11 所示。你可以看一下在公式栏出现的完整的函数表达式。

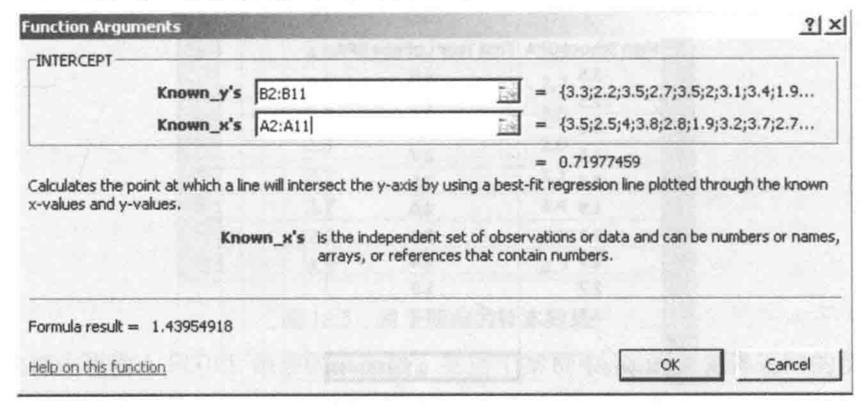


图 16.10 完成的 INTERCEPT 函数对话框

B14 =INTERCEPT(B2:B11,A2:A11)

	A	B	C
1	High School GPA	First Year College GPA	
2	3.5	3.3	
3	2.5	2.2	
4	4.0	3.5	
5	3.8	2.7	
6	2.8	3.5	
7	1.9	2.0	
8	3.2	3.1	
9	3.7	3.4	
10	2.7	1.9	
11	3.3	3.7	
12			
13	Slope (b)	0.7039	
14	Intercept (a)	0.7198	

图 16.11 完成的 INTERCEPT 函数

结果与我们手动计算的回归线的斜率(b)和截距(a)所得数值一样,我们使用 SLOPE 和 INTERCEPT 函数是做了相同的工作。现在,如果你想变得更有效率,我们可以建立一个如图 16.12 所示的工作薄,其中单元格 B18 包含一个公式,这个公式将两个函数组合起来,依据实际的 X 值来计算一个估计值(在这个案例中,X 的实际值是 3.5,在单元格 A2)。一定要在公式栏看看这个公式,以了解我们可以如何利用 Excel 的即有功能来做组合计算。非常神奇!

B18 fx =B13*A18+B14

	A	B
1	High School GPA	First Year College GPA
2	3.5	3.3
3	2.5	2.2
4	4.0	3.5
5	3.8	2.7
6	2.8	3.5
7	1.9	2.0
8	3.2	3.1
9	3.7	3.4
10	2.7	1.9
11	3.3	3.7
12		
13	Slope (b)	0.7039
14	Intercept (a)	0.7198
15		
16		
17	Actual Score	Predicted Score
18	3.5	3.1834

图 16.12 基于 SLOPE 函数和 INTERCEPT 函数进行估计

使用奇妙的分析工具计算回归方程

斜率、截距和最后的回归方程(你现在已经了解)涉及的计算非常多。但是 Excel 给我们提供了很多工具使得计算工作变得简单。我们使用分析工具库中的回归工具和之前在图 16.5 中所示的数据。

1.点击数据→数据分析→回归,你就会看到如图 16.13 所示的回归的数据分析对话框。

2.在 Y 值输入区域选框中,输入你想让 Excel 用于回归计算的因变量的数据范围。如图 16.5 所示,我们想分析的数据在单元格 B1 到 B11,包括数据列的标签。

3.在 X 值输入区域选框中,输入你想让 Excel 分析的自变量的数据范围。如图 16.5 所示,我们想分析的数据在单元格 A1 到 A11,包括数据列的标签。

4.点击标志选框。

5.在对话框的输出选项部分点击输出区域选项,输入你想让 Excel 将分析结果返回的单元格位置。在这个案例中,我们选择 D1。图 16.14 是完成的对话框呈现的形式。

6.点击确定。大家就可以看到图 16.15 所示的输出结果(我们进行了编辑,看起来很简明)。既简单又快速。

过程就是这样。我们现在得到了什么? 一些分析结果超出了我们本章讨论的范围。

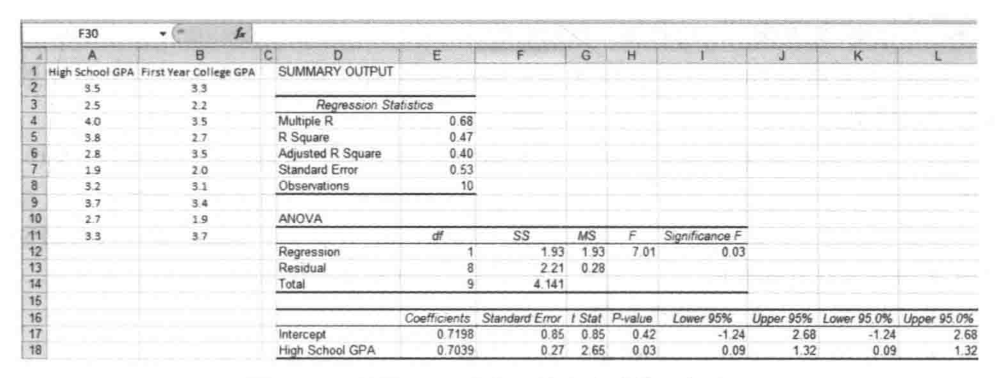
图 16.13　来自数据分析工具库的回归对话框

图 16.14　完成的回归对话框

但这不要紧，先让我们来看看如何讨论这些结果的意义。

在输出结果的底部你可以看到截距是 0.7198（与我们使用 INTERCEPT 函数所得结果一样），High School GPA（高中 GPA 成绩）是 0.7039（与我们使用 SLOPE 函数所得结果一样）。在这些值的右侧，有一栏是“P-value”，是对任何独立变量（高中 GPA 成绩）对于理解结果或者依赖变量（大学第一年 GPA 成绩）所做贡献的显著性的测量。在这个案例中，p 值是 0.03，因此我们保证估计变量可以很好地进行估计。

在输出结果的顶端是 multiple R（多元相关系数值）（0.68），在我们这个案例中只有一个估计变量，因此这个值就是皮尔森相关系数值。我们也发现 R value squared（决定系数值）（0.47）。我们在第 5 章已经学习过决定系数，即可以被另一个变量解释的变量的方差的大小。这个值很重要。



F30

	A	B	C	D	E	F	G	H	I	J	K	L
1	High School GPA	First Year College GPA		SUMMARY OUTPUT								
2	3.5	3.3										
3	2.5	2.2		Regression Statistics								
4	4.0	3.5		Multiple R	0.68							
5	3.8	2.7		R Square	0.47							
6	2.8	3.5		Adjusted R Square	0.40							
7	1.9	2.0		Standard Error	0.53							
8	3.2	3.1		Observations	10							
9	3.7	3.4										
10	2.7	1.9		ANOVA								
11	3.3	3.7			df	SS	MS	F	Significance F			
12				Regression	1	1.93	1.93	7.01	0.03			
13				Residual	8	2.21	0.28					
14				Total	9	4.141						
15												
16					Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17				Intercept	0.7198	0.85	0.85	0.42	-1.24	2.68	-1.24	2.68
18				High School GPA	0.7039	0.27	2.65	0.03	0.09	1.32	0.09	1.32

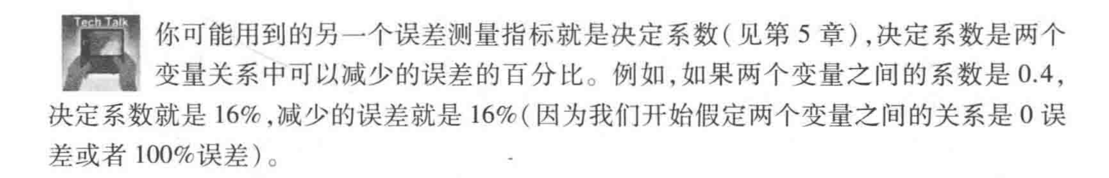
图 16.15 使用 Excel 分析工具库完成的回归分析

预测是否足够好?

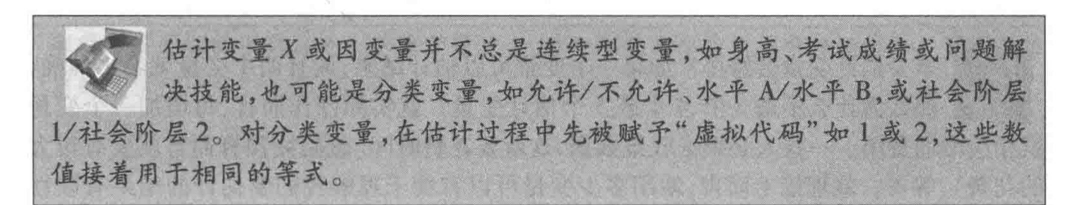
如何测量我们依据一个结果估计另一个结果的工作完成的好坏? 我们知道两个变量之间的相关系数的绝对值越大就能越好地进行估计。理论上说这没错。但是就实际估计而言,在我们首次计算回归线的公式时,我们就能够看到估计值(Y')和实际值(Y)之间的差异。

例如,如果回归线的公式是 $Y'=0.702(2.8)+0.726$,X 值为 2.8 的 Y 的估计值(Y')是 0.702(2.8)+0.726 或 2.69。我们知道对应 X 值的实际 Y 值是 3.5(见表 16.1 所示的数据集)。3.5 与 2.69 之间相差 0.81,也就是我们所说的**估计误差**(error of estimate)。

如果我们考虑所有的偏差,我们可以计算每一个数据点偏离于估计的数据点的平均数量,即**标准估计误差**(standard error of estimate)。这个值告诉我们估计的不准确性程度。就如你可以想到的,两个值之间相关程度越高(估计也越好),估计误差就越小。实际上,如果两个变量之间完全相关(+1 或-1),标准估计误差就是 0。为什么? 因为是完全估计,所有的实际数据点都落在回归线上,以 X 估计 Y 的过程中没有误差。



你可能用到的另一个误差测量指标就是决定系数(见第 5 章),决定系数是两个变量关系中可以减少的误差的百分比。例如,如果两个变量之间的系数是 0.4,决定系数就是 16%,减少的误差就是 16%(因为我们开始假定两个变量之间的关系是 0 误差或者 100%误差)。



估计变量 X 或因变量并不总是连续型变量,如身高、考试成绩或问题解决技能,也可能是分类变量,如允许/不允许、水平 A/水平 B,或社会阶层 1/社会阶层 2。对分类变量,在估计过程中先被赋予"虚拟代码"如 1 或 2,这些数值接着用于相同的等式。

估计变量越多就越好? 也许是

本章我们使用的所有案例都是一个结果变量和一个估计变量。有些情况下回归使用不止一个估计变量或独立变量估计一个特定的结果。如果一个变量能够以一定的准确性

估计一个结果，那么两个变量是否能更好地估计？

例如，如果高中 GPA 是大学 GPA 很好的指示器，那么高中 GPA 加上课外活动数量呢？回归模型从

$$Y' = bX + a \tag{16.8}$$

变成

$$Y' = b_1X_1 + b_2X_2 + a \tag{16.9}$$

其中：Y'表述被估计的数值，

X_1 表示第一个独立变量的数值，

X_2 表示第二个独立变量的数值，

b 表示对应变量的回归权重。

与你猜测的一样，这个模型就是**多元回归**（multiple regression）。因此，在理论上说就是以两个自变量而不是一个来估计结果。但是只有在某些条件下你才会想增加另外的自变量。首先，你增加的任何变量对于理解因变量有独特的贡献，否则就没有被使用的意义。我们所说的独特是什么含义？增加的变量需要解释第一个估计变量不能解释的被估计变量的差异。也就是说这两个变量综合起来可以比任何一个变量单独估计更好地估计 Y。

在我们的案例中，参与课外活动的水平可以做出独特的贡献。那我们是否可以增加一个变量如每个高中生的学习时间作为第三个独立变量或估计变量？因为学习时间与 GPA（我们的另一个估计变量）的相关程度很高，所以学习时间对于估计大学 GPA 不能提供更多信息。因此寻找另外的变量（例如推荐信的级别）可能比花费时间收集学习时间的数据更为有效。

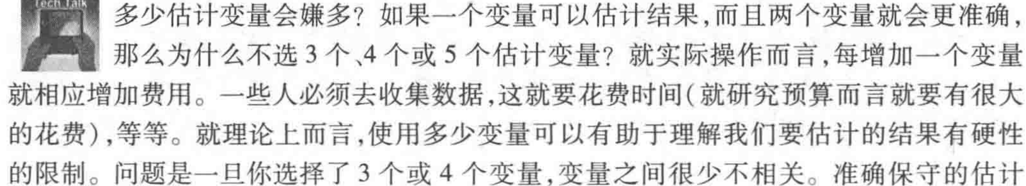
使用多元估计变量的重要原则

如果要使用不止一个估计变量，要谨记遵守下面的两项重要的原则：

1.如果选择一个自变量估计一个结果，要选择与被估计变量（Y）相关的估计变量（X）。也就是这两个变量有共变的部分（记住，它们应该相关）。

2.如果选择不止一个自变量或估计变量（如 X_1与 X_2），要尽量选择相互独立或者不相关的变量，但是都要与结果变量或被估计变量（Y）相关。

简而言之就是你只想要独立变量或估计变量与依赖变量相关，而彼此不相关。也就是每个独立变量尽可能在估计依赖变量或被估计变量时做出独特的贡献。



多少估计变量会嫌多？如果一个变量可以估计结果，而且两个变量就会更准确，那么为什么不选 3 个、4 个或 5 个估计变量？就实际操作而言，每增加一个变量就相应增加费用。一些人必须去收集数据，这就要花费时间（就研究预算而言就要有很大的花费），等等。就理论上而言，使用多少变量可以有助于理解我们要估计的结果有硬性的限制。问题是一旦你选择了 3 个或 4 个变量，变量之间很少不相关。准确保守的估计比包含太多变量、浪费太多资金和估计效果的估计更有优势。

小　结

估计是简单相关系数的特殊应用,而且是检验复杂关系的非常有力的工具。这可能使得本章比其他各章要难一些,但是你会受益于你的所学,特别是当你能够将此应用到你所阅读的研究报告和期刊文章中时。随着有关推论统计的大多数章节的结束,我们可以进一步将统计技术应用到样本规模非常小或者假定其数据分布不是正态分布的样本中。

练　习

1.第16章数据集1是一个群体参与定时测试的数据。数据是参与者完成每一个项目的平均时间(反应时间),以及项目正确的数量(正确数量)。

a.给出依据正确数量估计反应时间的回归等式。使用SLOPE函数和INTERCEPT函数建立这个等式。

b.如果正确数量是8,估计反应时间。

c.对每一个估计的反应时间来说实际的正确数量和估计的正确数量之间的偏差是多少?

2.贝特斯对估计有多少75岁的老人会患阿尔茨海默病感兴趣,并以教育水平和10级测量的一般健康状况作为估计变量。但她也对其他估计变量感兴趣。回答下面的问题。

a.在选择其他估计变量时应遵守什么标准,为什么?

b.定义其他两个可能与阿尔茨海默病发病相关的估计变量。

c.如果选择四个估计变量(教育水平、一般健康状况以及你定义的两个新的变量),给出可能的回归等式。

3.你对预测生意成功与否很感兴趣,你是学统计学的。为什么不出去收集15个不同的变量并进行变量的组合,来分析怎样的变量组合可以最好地预测成功(不论你想怎样定义和测量成功)?

4.去图书馆,在你感兴趣的领域的研究中选择三个不同的线性回归的案例。如果研究包含不止一个估计变量也可以。对应每个研究回答问题。

a.哪一个是自变量,哪一个是因变量?

b.如果有不止一个自变量,关于变量之间是否相互独立研究者进行了怎样的讨论?

c.三个研究中哪一个提供的表明依赖变量是依据自变量进行估计的证据最少,为什么?

5.现在应用本章提供的信息预测获得超级杯胜利者的机会。乔教练非常想知道一年中比赛胜利的平均次数是否可以预测超级杯的成绩(胜利或失败)。变量X是在过去10个赛季赛事胜利的平均次数。变量Y是过去10个赛季某个队是否获得超级杯。下面是给出的数据:

球　队	过去十年赛事胜利的平均次数	是否赢得超级杯（1=是，0=否）
Savannah Sharks	12	1
Pittsburgh Pelicans	11	0
Williamstown Warriors	15	0
Bennington Bruisers	12	1
Atlanta Angels	13	1
Trenton Terrors	16	0
Virginia Vipers	15	1
Charleston Crooners	9	0
Harrisburg Heathens	8	0
Eaton Energizers	12	1

a.如何评价赛事胜利的平均次数作为队伍是否赢得超级杯的估计变量的有效性？

b.使用分类变量（如 1 或 0）作为依赖变量的优点是什么？

c.其他的可用于估计因变量的变量有哪些？为什么选择这些变量？

6.为什么估计标准误越高估计变量的预测力就越低？

7.在第 16 章数据集 2 中你会发现三个变量的数值。结果变量是总的健康状况（Health），数值越高越健康；估计变量是对甜食的偏好（Sweets），较大的数值表示“喜欢甜食”；性别，1 是男性，2 是女性；体重指数（body mass index，BMI），肥胖程度测量指数，这个数值越高表明越胖。使用分析工具库中的回归工具分析这三个变量作为总的健康状况的估计变量的效果如何。

非正态分布时做什么——卡方和其他非参数检验 17

本章你会学到什么 ☺☺☺☺ ☒☒☒☒

- 什么是卡方检验,何时使用
- 对 CHISQ.TEST 函数的介绍
- 非参数统计的简要介绍以及何时、如何使用非参数统计

非参数统计的介绍

到现在为止,我们在《爱上统计学:Excel》这本书中介绍的每一种统计检验方法几乎都假定你所使用的数据集具有确定的特征。例如,支持均值(独立均值与不独立均值)之间 t 检验的一个假定就是每个群体的方差是齐性的,或者是类似的。而且这项假定可以检验。大多**参数统计**(parametric statistics)的另一项假定是样本要大到足以代表总体。统计学家已经发现样本规模达到 30 就可以满足这项假定。到现在我们已经学过的许多统计检验都是稳健的或者强有力的,哪怕这些假定中有一个不成立但是检验仍然有效。

但如果上述假定都不成立你将怎么做?最初的研究问题肯定只得继续提出并回答。这就是我们使用**非参数统计**(nonparametric statistics,也称为自由分布统计)的时候。这些统计检验方法不遵循相同的"规则"(也就是不需要和前面介绍的参数检验一样的假定),也同样很有价值。非参数检验的使用还可以让我们分析频数数据,例如不同年级的学生数量或获得社会保障的人口比例。

例如,如果你想知道最近的选举中赞同教育券的投票人数是否符合你的随机预测,或者是否真的存在偏好模型,那么我们就使用非参数统计也就是**卡方检验**(chi-square)。

在这一章我们介绍卡方检验——最常用的非参数检验之一,还将简要地介绍其他非参数检验方法,这样你就会对可能用到的一些非参数检验方法有一定的了解。

单样本卡方检验介绍

卡方是很好的非参数检验方法,可以用于确定你在频数分布中观察到的结果是否就是你随机预期的结果。单样本卡方只包括一维,就如你在这里看到的案例。双样本卡方检验包括两个维度,例如赞同教育券是否独立于政党背景和性别。这些都是后话(也许本书下一个版本会讲)。

例如,这里的数据是随机选自加利福尼亚州索诺玛郡 1990 年人口调查的一个样本数

据。就如你所看到的，这个表以不同教育水平划分数据。

教育水平			
没有上大学	上了大学	获得学位	总计
25	42	17	84

这里的问题是回答者的数量是否在各个教育水平等量分布。要回答这个问题就要计算卡方值（看起来是这个样子χ^2）并进行显著性检验。在这个案例中，卡方值等于 11.643，这个值在 0.05 显著水平下是显著的。结论就是这个案例中回答者在不同教育水平的分布不是等量的分布。换句话说，这不是随机出现的结果。

单样本卡方检验的原理是，就任何事件的发生而言都可以很容易地计算随机预期的结果。你可以通过事件总体发生的数量除以层级或者分类的数量来实现。在我们的人口调查样本中，观察到的事件发生的总数量是 84。我们随机预期 84/3 即 28 个（频数的总和 84 除以分类的总数 3）回答者分别落在教育水平的三个分类中。

现在我们来看我们随机预测的数量和实际观察的数量之间的差异程度。如果预期的数量和实际观察的数量没有区别，卡方值就等于 0。

接下来让我们具体地了解如何计算卡方值。

计算卡方检验统计量

卡方检验要进行观察值和随机预期值的比较。公式 17.1 就是单样本卡方检验的卡方值计算公式。

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{17.1}$$

其中：χ^2 表示卡方值，

$\sum$ 是求和符号，

O 表示观察频数，

E 表示预期频数。

下面是我们用于计算卡方值的数据。

对教育券的态度			
赞同	中立	反对	总计
23	17	50	90

下面是检验这个统计量的著名的八个步骤。

1.零假设和研究假设的表述。公式 17.2 所示就是零假设，表示每一分类内事件发生的频数或比例没有差异。

$$H_0: P_1 = P_2 = P_3 \tag{17.2}$$

零假设中 P 表示每一分类中事件发生的百分比。零假设表示分类 1（赞同）、分类 2（中立）和分类 3（反对）中个案所占的百分比相等。我们只使用三个分类，如果情况允许这个分类数字可以扩展，只要这些分类相互排斥，也就是任何一个观察值只能落在一个分

类中。例如，你不能同时既是男性又是女性。当然你也不能同时既赞同又反对教育券。

公式 17.3 所示就是研究假设，表示每个分类中事件发生的频数或百分比不同。

$$H_1:P_1 \neq P_2 \neq P_3 \tag{17.3}$$

2.设置零假设的风险水平(显著性水平，或第一类错误)。第一类错误设置在 0.05。

3.选择合适的检验统计量。相互区别的分类(例如赞同、中立和反对)的频数和百分比之间的任何检验需要使用卡方。我们之前使用的选择统计检验类型的流程图不适合非参数检验过程。

4.计算检验统计值(也称为实际值)。现在回到之前关于教育券案例的数据并建立一个数据表用于帮助我们计算卡方值。

分类	O(观察频数)	E(预期频数)	D(偏差)	$(O-E)^2$	$(O-E)^2/E$
赞同	23	30	7	49	1.63
中立	17	30	13	169	5.63
反对	50	30	20	400	13.33
总计	90	90			20.59

我们依据下面的步骤准备上面的数据表。



1—键入各个分类(Category)——赞同、中立和反对。要记住这三个分类相互排斥。任何数据点只能落在一个分类内。

2—键入观察频数(O)，表示实际收集到的数据。

3—键入预期频数(E)，是观察到的频数总和除以分类数量(3)，或者 90/3 = 30。

4—对应每一行(D)，就是观察频数减去预期频数。以预期频数减去观察频数也可以，因为这个值在下一步中将进行平方。

5—计算观察值和预期值差的平方。你可以在$(O-E)^2$一栏中看到这些值。

6—观察值和预期值偏差的平方除以预期的频数。你可以在$(O-E)^2$/E 一栏中看到这些值。

7—对最后一栏求和，你就会得到总的卡方值 20.59。

5.使用特定统计量的临界值分布表确定拒绝零假设需要的值。现在我们需要查阅附录 B 的表 B5，表 B5 列出了卡方检验的临界值。

我们的第一个任务是确定自由度(df)，自由度近似于数据表中的分类数量。对现在选定的统计检验量来说，自由度是 $c-1$，其中 c 等于行数，或 $3-1=2$。

使用这个数字(2)、你愿意承担的风险水平(之前定义的 0.05)，你可以使用卡方分布表查阅临界值。这个值是 5.99。因此，在显著水平为 0.05、自由度为 2 的情况下拒绝零假设需要的值是 5.99。

6.比较实际值和临界值。实际值是 20.6，拒绝零假设也就是在分类 1、分类 2 和分类 3 发生的频数相等需要的临界值是 5.99。

7.和 8.做出决定。

如果实际值大于临界值就不能接受零假设。如果实际值没有超过临界值，零假设就是最有力的解释。在这个案例中，实际值超过临界值——这个值足够大，我们可以说回答者在

三个分类中的分布不相等。实际上人们对教育券的选择上，赞成、中立或反对的频数分布存在差异。

Tech Talk 单样本卡方检验有一个常用的名称是拟合优度。这个名称暗含的问题是一个数据集“拟合”现存的数据集的程度。当然这个现存的数据“集”就是观察到的数据。“拟合”意味着存在另一个可以匹配观察到的数据的数据集。标准就是计算 χ^2 值的过程中计算的预期数据集。如果观察数据能拟合，就是观察数据和随机预期数据非常接近，而且不存在显著差异。如果观察数据不能拟合，就是观察到的数据和预期的数据不同。

如何解释 $\chi^2=20.6, p<0.05$

- χ^2 表示检验统计量。
- 20.6 是实际值，是使用本章之前给出的公式 17.1 计算所得的值。
- $p<0.05$（p 值是最重要的）表示对零假设的任何检验来说，投票频数在各个分类中随机等量分布的可能性小于 5%，因为我们以0.05作为研究假设比零假设更有解释力的标准，所以我们的结论就是两个数据集之间有显著差异。

使用 Excel 的 CHISQ.TEST 函数

Excel 不提供可以完成上面这些计算的函数。没有函数（或者分析包工具）可用于计算单样本的卡方值。但是 CHISQ.TEST 函数可以计算特定卡方值的概率，只要你输入自由度和卡方值。在我们之前的案例中，完成的函数类似于我们之前看到，返回的值就是3.363 31E-05，也就是 0.000 033 633 1（这个值非常小）。CHIDIST 只是提供给定自由度和统计量 χ^2 值的概率结果。一定要记住，如果你已经有了卡方值，使用这个函数时你不需要原始数据。只要将卡方值和自由度代入就可以，非常简单。

下面展示如何使用图 17.1 所示数据来进行计算。

1.选择你想输入 CHISQ.TEST 函数的单元格。在这里案例中，我们选择将 CHISQ.TEST 值返回到单元格 B6（选择这个位置没有特任何特别的原因）。

2.现在点击［公式］→［其他函数］→［统计］菜单选项，在下拉选项中选择 CHISQ.TEST 函数，你可以看到图 17.2 所示对话框。

	A	B	C
1	Category	Observed Frequency	Expected Frequency
2	For	23	30
3	Maybe	17	30
4	Against	50	30

图 17.1　卡方分析数据

3.在对话框的 Actual_range 选项框输入实际值的范围（从单元格 B2 到 B4）。

4.在图 17.3 所示对话框的 Expected_range 选项框中输入期望值的范围（从单元格 C2 到 C4）。

5.点击［确定］，不同分类等量分布的概率返回到单元格 B6，在图 17.4 中你可以看到具体的值是 3.363 31E-05（或 0.000 033 633 1）。这个值非常非常小。我们的结论是什么？投票模式肯定不是随机的。仔细看一下公式栏中公式的参数或者命令。

Function Arguments

CHISQ.TEST

Actual_range　= array

Expected_range　= array

=

Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

Actual_range　is the range of data that contains observations to test against expected values.

Formula result =

Help on this function　OK　Cancel

图 17.2　CHISQ.TEST 函数对话框

Function Arguments

CHISQ.TEST

Actual_range　B2:B4　= {23;17;50}

Expected_range　C2:C4　= {30;30;30}

= 3.36331E-05

Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

Expected_range　is the range of data that contains the ratio of the product of row totals and column totals to the grand total.

Formula result = 3.36331E-05

Help on this function　OK　Cancel

图 17.3　完成的 CHISQ.TEST 函数对话框

B6　=CHISQ.TEST(B2:B4,C2:C4)

	A	B	C	D
1	Category	Observed Frequency	Expected Frequency	
2	For	23	30	
3	Maybe	17	30	
4	Against	50	30	
5				
6		3.36331E-05		

图 17.4　CHISQ.TEST 函数值

实际上自己建立计算卡方值的公式也很容易,就如图 17.5 所示。我们输入相同的数据,这些数据在本章已经多次使用,接着你在公式栏中能看见正确的公式。当然这个公式是合适变量有三个层级的要求,因此如果你有更多的层级,公式就会更复杂一些。

C7	=((B3-C3)^2/C3)+((B4-C4)^2/C4)+((B5-C5)^2/C5)					
	A	B	C	D	E	F
2						
3	Republican	23	30			
4	Democrat	17	30			
5	Independent	50	30			
6						
7		Chi Square	20.6			

图 17.5 建立计算卡方值的公式



你刚才所学的卡方检验是单样本卡方或者拟合优度检验。另一个非常有用的检验类型是独立性检验。独立性检验比较两个因素在不同的分类中的发生频数是否相互独立。例如，你可能有类似如下的数据：

	课程及格	课程不及格
参与课外活动	22	18
没有参与课外活动	18	22

这组数据可以让我们判断参与课外活动是否对课程成绩有影响(即两个变量是否独立)。与单样本卡方检验类似，有了实际值和预期值就可以计算卡方值，其所对应的概率值就是这两个因素，相互独立的可能性。在这个案例中，返回的概率值是.37(我们使用CHISQ.TEST函数计算)，意味着这两个维度相互独立，即是否参与课外活动不影响课程成绩。

函数库

D9	=CHISQ.TEST(C2:D3,C6:D7)					
	A	B	C	D	E	F
1	实际值		及格	不及格		
2		参与课外活动	22	18		
3		没有参与课外活动	18	22		
4						
5	预测值		及格	不及格		
6		参与课外活动	20	20		
7		没有参与课外活动	20	20		
8						
9				0.371093		
10						
11						

图 17.6 使用 CHISQ.TEST 函数计算概率值

其他需要了解的非参数检验

你可能永远不需要非参数检验来回答你所提出的研究问题。但是，你也可能发现你分析的样本非常小(至少小于30)或者数据不满足支持参数检验的一些重要的假定。

Tech Talk

实际上，你使用非参数统计的最直接原因是你所研究的变量因所处的测量水平。我们在下章会进行更多的讨论，但是就现在而言，大多数数据是分类的或者是放在不同的分类中的(例如鲨鱼种类和喷气式飞机的种类)，或者是定序或等级排序的(如第

1、第2和第3),它们都需要表17.1中给出的某种非参数检验。

如果是这种情况,可以依据样本规模选择非参数统计。表17.1给出了你需要了解的一些非参数检验的内容,包括名称、使用目的和每一个方法应用的研究问题。要记住的是这个表中只给出了你可能用到的各不相同的众多检验方法中的一小部分。

表17.1 分析分类和定序数据的非参数检验

检验方法名称	何时使用	研究问题示例
分析分类数据		
检验变化显著性的McNemar检验	检验变化的"前和后"	打电话给对某个特定问题没有做出投票决定的候选人的效果如何?
Fisher精确检验	计算2×2列联表中每个结果的准确概率	投掷6枚硬币得到6个正面的准确概率?
单样本卡方检验(就是本章讨论的重点)	确定不同分类的事件发生次数是随机分布的	在最近的销售中品牌Fruities、Whammies和Zippes的销售量是否相同?
分析定序数据		
Kolmogorov-Smirnov检验	分析一个样本的数据是否来自特定的总体	一群小学生选择到哪一所小学去的判断是否具有代表性?
符号检验或中位数检验	用于比较两个样本的中位数	投票给候选人A的群体的收入中位数是否高于投票给候选人B的群体的收入中位数?
Mann-Whitney U检验	用于比较两个独立样本	群体A是否比群体B的学习转移速度快?——以正确数量测量
Wilcoxon等级检验	比较两个群体的差异的大小和方向	在帮助儿童语言技能的发展方面学前教育带来的效果是否是没有学前教育的两倍?
Kruskal-Wallis一元方差分析	比较两个或多个独立样本的总体差异	四个地区办公室管理者的等级差异如何?
Friedman(弗里德曼)二元方差分析	比较两个或多个独立样本在不止一个维度上的总体差异	四个地区办公室管理者的性别和等级差异如何?
Spearman(斯皮尔曼)等级相关系数	计算等级相关系数	高中最后一年的名次和大学第一年的名次之间的相关系数是多少?

小结

卡方检验是林林总总的非参数检验中的一种,可以帮助你回答不符合正态分布基本假定或者数据规模太小的统计检验问题。这些非参数检验是非常有价值的工具,即使现在提供的介绍非常有限,你还是会得到一些帮助。

练 习

1.使用下面的数据检验问题——最近的选举中民主党、共和党和无党派人士的投票人数相同。在 0.05 的显著水平下检验假设。

政治背景		
共和党	民主党	无党派人士
800	700	900

2.使用下面的数据在 0.01 显著水平下检验问题——男孩和女孩参加初级足球培训的人数相同。你的结论是什么？

性别	
男孩	女孩
45	55

3.下面的四个研究问题哪一个适合卡方检验？

a.两个数学班的平均成绩的差异。

b.1 班通过数学考试的人数和 2 班通过数学考试的人数不同。

c.今年通过碰撞检验的汽车数量与去年通过的数量不同。

d.比较橄榄球运动员与足球运动员的 100 码跑步速度。

4.在糖尿病研究所，病人通常被分为两类：一类是按规定服药的，一类是不按规定服药的。新的项目管理者有一个新的想法，让一些病人持续地监测血糖而告诉另外一些病人不需要监测血糖，并考察他们是否会因此在饮食控制方面有不同。监测血糖病人对饮食控制成功者和不成功者人数如下表。请问这个新项目是否有效？

成功	不成功
60	38

你应该了解的其他重要的统计过程 18

本章你会学到什么 ☺☺☺☺ ☒

- 高级统计过程综述,以及何时如何使用
- 你可能想学习的其他函数和数据分析工具

《爱上统计学:Excel》全书只是覆盖了整个统计学的一小部分。我们没有提供过多的知识,而是保持了基础级知识的简单、直接。

但是这并不意味着你阅读最近的研究文章或者在课堂讨论时不会碰到其他的分析技术,这些技术很重要,需要你有所了解。因此,为了让你所学更加全面,这里列出了其他一些统计技术,介绍它们的应用,并给出使用这些技术回答问题的研究案例。

要注意新的技术和工具始终在不断发展,对于分析大型数据非常有用,而且如果你有时间,你应该在课业中尽量去学习更高级的统计课程。但是就目前来说,简要的汇总性介绍可以扩展你的知识,让你至少可以在更广泛的统计学中先找到起点。

事后比较分析

统计学家在 F 检验方面有很大的进展(是否还记得方差分析,如果不记得,复习第 13 章)。F 检验是对两个或两个以上样本均值的稳健检验。在实践中这意味着你不能区别是群体 1 与 2、群体 1 与 3 还是群体 2 与 3(或者其他组合)之间的差异是导致总差异的原因。但是如果你非常聪明,你会知道事后比较分析,或知道可以进行事后分析比较的工具。这类方法有好几种,其中之一称为 Tukey's HSD (Honestly Significant Differences,真实的显著性差异)。

例如,蒙大拿大学的林恩·桑福德·凯斯特(Lynne Sanford Koester)和伊芙·拉赫蒂-哈珀(Eve Lahti-Harpee)研究父母如何"指导婴幼儿情绪调节、语言习得和参与社会交往"。他们比较听障和听力正常的父母-婴幼儿研究组在婴幼儿 6、9、12 和 18 个月的直觉教养行为水平。在分析中他们使用 Tukey's HSD 事后分析工具检验不同的父母-婴幼儿研究组之间的差异。

是否想了解更多?在线查找凯斯科(Koester, L. S.)和拉赫蒂-哈珀(Lahti-Harper, E) 2010 年发表在《美国听障研究年刊》(*American Annals of the Deaf*)第 155(1)期第 5-18 页的文章"新生儿 18 个月期间母亲-婴幼儿听力状况与直觉教养行为(Mother-infant hearing status and intuitive parenting behaviors during the first 18 months)"。

多元方差分析

方差分析（ANOVA）有多种应用的方式，每一种形式的设计都适合“两个以上群体的平均值比较”的特定情况。其中之一是多元方差分析（MANOVA），用于不止一个因变量的情况。也就是不只使用一个结果变量或因变量。如果因变量或结果变量之间相关（这种情况很常见，见第 13 章关于多元 t 检验的专业讨论），就很难确定处理变量对单个结果变量的影响。因此，需要多元方差分析来解决这个问题。

例如，印第安纳州立大学的乔纳森·普吕克（Jonathan Plucker）研究天才少年处理学校压力的性别、种族和年级差异。他使用的多元方差分析是 2（性别：男性和女性）×4（种族：高加索人、非裔美国人、亚裔美国人和西班牙人）×5（年级：从 8 年级到 12 年级）MANOVA。分析的多元变量是青少年应对量表的 5 个子表。使用多元技术可以相互独立地估计独立变量（性别、种族和年级）对 5 个子量表中每一个量表的影响。

是否想了解更多？查阅文献：查阅普吕克（Plucker, J.A.）1998 年发表在《天才教育杂志》（*Journal for the Education of the Gifted*）第 21 期 423-436 页的文章“天才青少年的应对策略的性别、种族和年级差异（Gender, race, and grade differences in gifted adolescents' coping strategies）”。

重复测量的方差分析

这是方差分析的另一种形式。重复测量的方差分析非常类似于其他的方差分析，你可以回顾第 13 章的两个或两个以上群体均值的差异检验。就重复测量的方差分析来说，参与者在一个要素上要测试两次。这也是称为“重复”的原因，你要对相同的因素在不同的时间点进行重复测量。

例如，B.伦迪、T.菲尔德、C.麦克布莱德、T.菲尔德与 S.拉尔吉（Brenda Lundy, Tiffany Field, Cami McBride, Tory Field, and Shay Largie）使用高中三年级和四年级的数据检验同性和异性最好朋友之间的相互影响。他们的主要分析之一是三个因素的方差分析：性别（男性或女性）、友谊（同性或异性）以及在高中的年级（三年级或四年级）。重复测量的要素是年级，因为是在不同年级重复测量过程。

是否想了解更多？查阅文献：B.伦迪和 T.菲尔德等人（Lundy, B., Field, T., McBride, C., Field, T., & Largie, S.）发表在 1998 年《青少年》（*Adolescence*）第 33 卷 130 期 279-289 页的文章“高中三年级和四年级中同性和异性最好朋友之间的相互影响（Same-sex and opposite-sex best friend interactions among high school juniors and seniors）”。

与 Excel 相关的内容是什么？分析工具库中的方差分析：可重复双因素分析。

协方差分析

这是我们介绍的最后一种方差分析。协方差分析（ANOVA）是特别有趣的一种形式，因为它允许你将群体之间最初的差异等量化。我们假定你赞助一个提高速度的项目，而且你想比较两组运动员在 100 码冲刺中能够跑多快。因为力量通常和速度有关，你必须做一些修正，这样力量就不能解释项目结束时的差异了。同时，你想了解去除力量因素之后的培训效果。你应该在培训项目开始之前测量培训者的力量，然后使用协方差分析来

调整基于最初力量的最后速度。

麦吉尔大学的米夏埃拉・希涅、约翰・林登与阿里・塔达什(Michaela Hyine, John Lyndon & Ali Tardash)在关于亲密行为和承诺对婚前性行为的接受程度和使用避孕物品的影响的研究中使用了协方差分析。他们使用协方差分析以社会接受度作为因变量(在其中寻找群体差异)、特殊事件的排序作为协变量。协方差分析能够保证社会接受度的差异将会以排序进行修正,这就会成为受到控制的一种差异。

若想了解更多,请查阅:希涅、林登与塔达什(Hyine, M., Lyndon, J. E. & Tardash, A.)发表在 1997 年《妇女心理学季刊》(*Psychology of Women Quarterly*)第 21 期 447-464 页的文章"承诺、亲密行为以及女性对婚前性行为和避孕物品的准备的认识(Commitment, intimacy, and women's perceptions of premarital sex and contraceptive readiness)"。

多元回归

你已经在第 16 章学习了如何利用一个变量的值来估计另一个变量的值。通常社会科学和行为科学的研究者都使用不止一个变量来估计另一个变量。我们在第 5 章和第 16 章都有所涉及,这里更多地介绍多元回归。

例如,我们可以很容易假定父母读写方面的行为(如家里有很多书)与他们子女读书的多少和好坏有关系。那么研究父母的年龄、教育水平、读写活动以及与子女共同阅读这些变量对子女的早期语言能力、读书兴趣的影响一定十分有趣。保拉・吕蒂宁、玛丽亚-莱纳・拉克索与安娜-迈亚・波伊凯乌斯(Paula Lyytinen, Marja-Leena Laakso, & Anna-Maija Poikkeus)进行了这方面的研究,并使用逐步回归分析来检验父母背景变量对孩子的读写能力的影响。他们发现母亲的读写活动和教育水平能够显著地影响子女的语言能力,而母亲的年龄和共同阅读没有显著影响。

是否想了解更多? 查阅文献:吕蒂宁、拉克索与波伊凯乌斯(Lyytinen, P., Laakso, M-L. & Poikkeus, A-M.)1998 年发表在《欧洲教育心理学杂志》(*European Journal of Psychology of Education*)第 3 期 297-308 页的文章"父母对子女的早期语言和读书兴趣的影响(Parental contributions to child's early language and interest in books.)"。

Logistic 回归

Logistic 回归与我们刚提到的多元回归有些不同。基本上说,Logistic 回归的结果变量或者因变量是一个分类变量,类似"喜欢"或者"不喜欢",而且预测变量可能是连续变量,例如分值是 1 到 100 的考试成绩,或者是分类变量,例如成员身份(属于或者不属于)。这样就形成两种类型——二元逻辑回归,也就是只有两个结果(如"合格的""不合格的");多元逻辑回归,也就是需要预测的变量有不止两个分类。

例如,斯蒂芬・什曼谢科(Stephen Shmanske)研究决定高尔夫运动员选择加入哪一个联赛的因素。他分析奖金金额、竞争力量、运动员技能这几个因素。他发现奖金的数量是最重要的因素,其次是技能。

是否想了解更多? 查阅文献,什曼谢科(Shmanske, S.)2009 年发表在《运动财经国际杂志》(*International Journal of Sport Finance*)第 4(2)期 114-135 页的文章"高尔夫竞赛:PGA Tour(巡回赛球员协会)高尔夫职业选手进入联赛的选择(Golf match: The choice by PGA Tour golfers of which tournaments to enter.)"。

如何使用 Excel？你同样可以使用分析工具库中的回归工具。

因子分析

因子分析是基于不同项目的彼此相关程度而形成因子或聚类的技术。每一个因子代表几个不同的变量，在特定的研究中用因子表示结果比单个变量更有效。在使用这项技术的过程中，目标是用更综合的名称如一个因子来描绘彼此相关的项目。而且用于描述变量群的因子的名称不能是随心所欲的——名称要反映内容和概念的彼此相关。

例如，西安大略大学的戴维・沃尔夫(David Wolfe)和他的同事尝试着去了解 12 岁以前发生的被虐待的经历对青少年时期的同伴关系和交往关系的影响程度。为了完成这项研究，研究者收集了多个变量的数据而且分析了所有变量之间的关系。那些似乎包含彼此相关项目的变量被认定为因子，如这项研究中名称为虐待/责备因子。另一个因子的名称是积极的沟通，由 10 个不同的项目构成，这些项目之间都彼此相关。

是否想了解更多？查阅文献：沃尔夫、韦克勒、莱策尔-贾菲与勒费弗(Wolfe, D. A., Wekerle, C., Reitzel-Jaffe, D., & Lefebvre, L.)1968 年发表在《发展心理学》(*Development and Psychopathology*)第 10 期 61-85 页的文章“受到虐待和没有受到虐待的年轻人之间与消极关系相关的因子(Factors associated with abusive relationships among maltreated and non-maltreated youth)”。

数据挖掘

数据挖掘严格来说不是检验关系或者发展趋势的统计技术，而是处理大数据的工具。大数据已经成为社会和行为科学中不可或缺的存在。就其他领域来说，数据挖掘是寻找数据集中模式的重要工具。数据挖掘首先被商业团体用于发现财经发展趋势，但是现在已经被心理学家、教育学家、护士以及其他人员应用。数据挖掘本质上非常有解释力，有时是主要分析方法的辅助，而且也用于探索来自不同数据集的变量之间的关系。数据挖掘有时也指数据或者知识探索。就如你所预期的，随着这项技术的流行，新的软件工具已经开发出来，能很好地进行数据挖掘(这在计算机发明之前是不可能的)。

例如，对年轻一代的学生来说，你们知道在线购物行为在过去的十年迅猛发展。帕米拉・诺鲁姆(Pamela Norum)的一项研究是分析大学生的在线购买行为，她使用数据挖掘技术考察 4 688 个大学生在 9 类商品方面的购买行为。什么因素在在线购买(主要指大学生)中扮演关键角色？备选的因素是年龄、性别、收入、汽车拥有情况、确定网站安全的能力，以及强迫性购买行为。

是否想了解更多？查阅资料：诺鲁姆(Norum, P. S.)2008 年发表在《家庭和消费科学研究杂志》(*Family and Consumer Sciences Research Journal*)第 36(4)第 373-388 页的“学生网络购买(Student Internet purchases)”。想购物？我的手机在哪儿？

路径分析

这是分析相关的另一项统计技术，但是在分析因素之间的关系时允许一定意义上确定相关的方向和因果关系。路径分析基本上通过变量间关系的理论假定来分析关系的方向，接着检验关系的方向是否得到数据的支持。

例如,埃夫克里德、帕帕扎基、帕帕托尼奥斯与基奥赛格鲁(Efklides, Maria Papadaki, Georgia Papantoniou, & Gregoris Kiosseoglou)研究个体学习数学过程中对难度的感知。为完成这项研究,他们进行了几项不同类型的测试(如认知领域的测试)而且发现对难度的感知主要受到认知(问题解决)因素而不是情感(情绪)因素的影响。路径分析最有趣的应用之一是结构方程技术,它是以图形展示所有考虑到的不同因素之间的关系。应用这项技术,你可以看到某个变量和某个变量之间相关以及相关的强度。接着你就可以判断数据与模型之间的适合程度。

是否想了解更多?查阅文献:埃夫克里德、帕帕扎基、帕帕托尼奥斯与基奥赛格鲁(Efklides, A., Papadaki, M., Papantoniou, G., & Kiosseoglou, G.)1998 年发表在《欧洲教育心理学杂志》(*European Journal of Psychology of Education*)第 2 期 207-226 页的文章"个体困难度感知差异:以学校数学课为案例(Individual differences in feelings of difficulty: The case of school mathematics)"。

结构方程模型

结构方程模型(SEM)是比较新的技术,自从 1960 年代早期引入以来已经变得十分流行。一些研究者觉得这项技术是回归、因子分析和路径分析的统称。其他的研究者相信这项技术本身代表着完全不同的方法,是基于变量之间关系的方法(类似于之前我们描述的三项技术)。

结构方程模型和其他高级统计技术如因子分析的主要差别是结构方程模型是实证的,而不是解释的。换句话说,研究者更多地使用结构方程模型来确定已经提出的模型是否发挥功效(也就是数据适合模型)。解释性技术用于发现特定的关系,很少(但不是没有)预先进行模型建构。

例如,希瑟·约塔姆、肯尼斯·舍尔与菲利浦·伍德(Heather Gotham, Kenneth Sher, & Phillip Wood)研究年轻人的酒精饮用紊乱、成年前的变量(性别、家庭酒精饮用历史、儿童期的压力、高中班级的排序、宗教背景、神经过敏症、个性外向、精神病史)和年轻人的发展任务(完成学位、全职工作、婚姻)之间的关系。他们使用结构方程模型技术发现成年前的变量相比年轻人的发展任务更为突出地导致年轻人酒精饮用紊乱。

是否想了解更多?查阅文献:约塔姆、舍尔与伍德(Gotham, H.J., Sher, K.J., & Wood, P.K.)2003 年发表在《酒精研究杂志》(*Journal of Studies on Alcohol*)第 64(1)期 32-34 页的文章"青年期的酒精依赖和发展任务的完成(Alcohol involvement and developmental task completion during young adulthood)"。

小　结

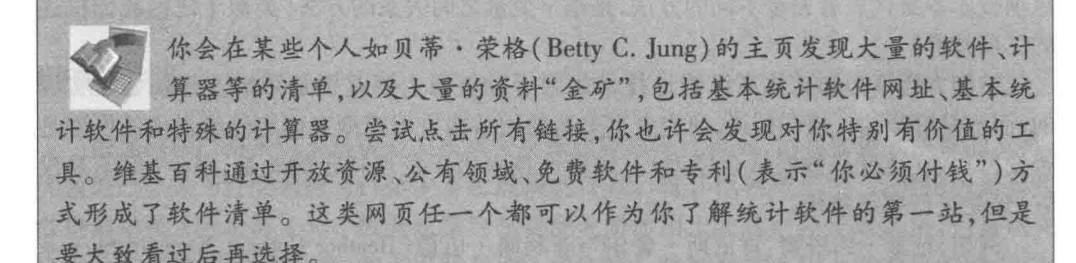
即使你近期不会使用这些高级的统计过程,但是你至少需要对它们有所了解,因为你肯定会在不同的研究出版物中看到这些技术被提到,或者在你选修的其他课程中被提到。随着对基本的统计技术(本书到现在为止的所有章节)的了解,你可以确信你已经掌握了大量的基础统计学知识。

19 统计软件简介

本章你会学到什么 ☺☺☺☺☺

- 介绍可用于分析、绘图以及更好地理解数据的各种类型的统计软件

在你努力学习和使用基础统计时不需要让自己成为书呆子,也就是不需要了解并深入到所有各种可用的计算机软件中。本章的目的是大概介绍一些常用的统计软件,包括它们的特征以及用法。但是在进行具体描述之前,先给出几个建议。



你会在某些个人如贝蒂·荣格(Betty C. Jung)的主页发现大量的软件、计算器等的清单,以及大量的资料"金矿",包括基本统计软件网址、基本统计软件和特殊的计算器。尝试点击所有链接,你也许会发现对你特别有价值的工具。维基百科通过开放资源、公有领域、免费软件和专利(表示"你必须付钱")方式形成了软件清单。这类网页任一个都可以作为你了解统计软件的第一站,但是要大致看过后再选择。

选择合适的统计软件

这里给出的建议经得起时间的检验,可以确保你得到你想要的统计软件。

1.不论统计软件的价格高昂(如 SPSS)还是低廉(类似 EcStatic),在你购买之前一定要试用。上述清单中的统计软件几乎都提供了试用版(通常在官方网站给出),你可以下载,而且在一些情况下你甚至可以请他们给你邮寄试用版本的软盘或者 CD。这些版本通常具备全部功能而且可以持续使用 30 天,给予你足够的时间在购买之前试用。

2.既然我们刚才提到价格,直接从制造商购买软件可能是最昂贵的方式,特别是你直接购买没有要求学生或教师折扣(有时他们把这称为教育折扣)。你学校的书店可能给你一个折扣,而邮购公司可能给你更低的折扣(再一次要求教育折扣)。你可以在大众计算机杂志中找到这些销售商的免费电话。记住,软件分析格总是在变化。

3.许多编制统计分析软件的销售商提供两种版本:一种是商业版本;一种是学术版本。他们在内容上通常都一样但是价格有差异(有时非常大),如果你想要学术版本,要确定这个版本和商业版本完全相同,如果不相同,你就要问自己是否可以忍受这种差异。为什么学术版本这么便宜?如果你是学生,公司希望等你毕业之后,你就会进入一些薪水高的公

司并购买完全的版本。

4.在开始之前很难准确地知道需要什么,但是一些软件包以模块的形式出现,你不需要购买全部便可得到你工作需要的统计工具。可以阅读公司的宣传手册,也可打电话咨询。

5.共享软件是另一个选择,而且有许多这样的共享软件。共享软件是分割软件的方法,这样就可以购买你想要的部分。共享软件价格几乎都很合理,通常比商业版本好;而且,如果你已经付费,你可以帮助聪明的作者继续努力开发比现在的版本更好的软件。

6.不要购买任何不提供电话技术支持的软件,或者至少要有类似电子邮件的联系方式。这很好验证,拨打技术支持电话(在你购买之前)来确定等待多长时间他们才接电话。如果你等待了 20 分钟,这意味着他们不能提供足够的技术支持来快速回答使用者的问题。或者你给他们发邮件却没有得到回复,那么就考虑其他软件。

7.几乎所有的大型统计软件包都有相同的功用——差别在于完成的方式。例如,SPSS、Minitab 与 JMP 在数据分析方面都做得很好,而且很受欢迎。但是很小的不同也可能是很大的差异。例如 Minitab 可以同时打开两个数据文件而 SPSS 却不可以。亲自验证一下。

8.确定你的硬件可以运行你要使用的软件。例如,大多数软件不受限于你要分析的个案和变量的数量,唯一的限制通常是你要用于储存数据文件的硬件驱动的大小。如果你的电脑运行速度很慢(不是奔腾系列),那么你就可能要等待很久,在 CPU 缓慢运行程序时眼巴巴看着时间流逝。在下载试用版本之前要确定硬件可以运行你要用的软件。同样要确定你的硬件可以支持 Mac 操作系统或 Windows 操作系统,并且可以实现兼容。

大概有哪些软件

可以得到的统计软件(超过 200 种)比你需要的多。下面列出一些最流行、最有特色的软件。记住许多软件功用相同。就如之前强调的,要尽可能在购买前试用。可以得到的软件很多,有的永久免费,有的提供免费的试用版本,有的要付费。根据自己的方便试用这些软件,直到你找到合适的。我们会介绍一些免费和开放软件(也就是会得到软件的编码,你可以改变或者提高软件功能),接着介绍一些一定程度上需要付费的软件。

学习本书的朋友可能更喜欢下面的介绍。

首先,免费和开放的那些

免费的并不差,在一些情况下,甚至就是你需要的。先了解一下哪些是免费的,可以先查阅免费统计软件。记住这里列出了"Completely Free"(完全免费)部分(其后没有附加说明),但是也不要忽略"Free, but..."(免费的,但是……)部分(其后有些如"学生版本""使用版本"等附加说明)。

Gnumeric 电子表格

排在我们清单第一位的是 Gnumeric 电子表格,执行的是通用公共许可(GPL)。这意味着这个软件是免费和开放的资源,你可以从不同资源处下载。这个类似 Excel(从形式和功能)的电子表经过 GNOME 集团的大量努力产生,这个集团花费大量的时间和资源建立了向每个人开放的软件。不像 Excel? 说到 Gnumeric,几乎本书所学的内容都可以应

用。甚至可以更好的应用，特别是对 Mac 系统爱好者来说（记住 Excel 最新 Mac 版本不支持数据分析工具库），Gnumeric 包含了你可能需要的所有 Windows 版本数据分析工具库的所有工具。

需要更多的信息？请查阅官方网站。

SSP（史密斯统计软件包）

SSP 是一款简单的统计软件包，可用于 Mac 和 Windows 操作系统，具备 SPSS 类型的很多功能，如计算基本的统计汇总、绘制图表、比较均值，进行方差分析、卡方检验和回归。你可以在图 19.1 看到简单 t 检验的结果。这是一款真的免费的软件包，很简单，而且可以满足你的需要。但是，Mac 用户要注意。Mac SO 最新的 Lion 不支持 SSP。

需要更多的信息？请查阅官方网站。

图 19.1　SSP 完成的简单 t 检验

VassarStats

"非常好"是我们对这个免费网址的评价，这个网址可以处理很多统计主题，如概率、频数、相关和回归、方差分析等。VassarStats 也包括很多的应用软件，如统计表格计算器、简单图表制作器。这个网址几乎可以处理本书描述的 Excel 的所有统计任务。

需要更多的信息？查阅官方网站。

R

R 是现在正发展的非常流行的开放资源软件包之一，最初是由贝尔实验室的人员开发。好消息是这个软件是你需要的全部，坏消息是开始使用这个软件需要花费些工夫，而且学习曲线有一些陡峭。如果你是初学者，能够和熟悉 R 的人一起学习，那么这个软件是最好学的。一旦你了解了这个软件，随着你对这个强有力的分析工具的知识的增长，你的

工作速度也会加快。

需要更多信息，请查阅官方网站。这里有很多相关的网址，你可以获得关于 R 的信息。

PSPP

PSPP 是 SPSS 软件（非常成功也非常贵）的一个开放资源变体。虽然安装过程有些复杂，但是整体上是一个非常复杂的、完整的、界面非常容易使用的软件分析包。

需要更多的信息？请查阅官方网站。

付费软件

好吧，免费软件已经介绍了很多。不要忘记先了解免费软件能做什么，如果还是不够用，再接着往下看这里的介绍。

StatCrunch（过去称为 WebSTAT）

这个软件的最大优点是什么？大概是 12 美元一年或者是 5 美元半年，而且最重要的是这个软件有网站支持。不需要下载软件——只需要激活，输入你的数据，然后按照你的要求运行。而且因为是网址支持，你可以在任何地方使用这个软件（和你的数据）。令人印象非常深刻，方便而且十分有趣。现在这个软件归培生教育集团（Pearson Education）所有和管理。

需要更多信息，请查阅官方网站。

JMP

JMP（现在是第 9 版）如广告上所说是“统计发现软件”。这个软件可以在 Windows、Mac 和 Linux 平台运行，而且是“将统计和图形结合在一起分析、理解和视觉化数据”的软件。JMP 的一个特点是用图形展示每一个统计分析结果，这样你总是可以看到数字和图形展示的两类分析结果。而且这是自动完成的，不需要你进行操作。

需要更多的信息，请查阅官方网站。

价格：学术版本是 595 美元。

Minitab

这是第一批可用于个人电脑的软件之一，而且现在已经是第 16 版（已经存在一段时间）了，这意味着可以看到这些年为了回应使用者的需求而发生的变化。新版本的一些显著特征如下：

- 可定制的菜单和工具栏；
- StatGuide™帮助解释结果；
- ReportPad™是报告产生器；
- 在线指导；
- 用户可编辑的说明；
- 新的图形产生器；
- 大量新的多元分析工具。

需要更多的信息，请查阅官方网站。

价格：完整版本是 1 395 美元，学术版本是 100 美元（也提供有时间限制的更便宜的版本）。

STATISTICA

StatSoft 提供了用于 Windows 操作平台的 STATISTICA 10 产品的集合。这个强有力的软件的优点是：自我产生对话框和模板（点击 OK，STATISTICA 会告诉你要输入什么）；可定制的界面；很容易和其他软件整合；STATISTICA Visual Basic 可以允许你实现 11 000 多种功能，而且可以依据这个发展环境设置新的功能；有使用宏命令实现任务自动化的能力。

网站的一个优点是提供电子版 STATISTICA 教科书（简称 EST），你可以完整下载（要有耐心，因为可能需要 30 分钟，这取决于你的网络连接速度），或者购买纸质版，需要 80 美元。

需要更多的信息，请查阅官方网站。

价格：每年 50 美元。

SPSS

SPSS 是现在使用的最流行的统计软件包之一。SPSS 有许多不同的模块，包含了统计分析的所有方面，包括基本统计和高级统计，而且一个版本可以用于几个操作平台（Windows XP/Vista/7, Mac OS X [Tiger, Leopard and Lion], and Linux）

SPSS 20 版的更新优点之一是很容易使用的新图形类型和模板。图 19.2 所示是 SPSS 建立的简单的条形图，可以告诉你一个群体中男性和女性的数量。SPSS 还有一个特点是强大的报告编辑功能。SPSS 现在归 IBM 所有，因此好消息是有大量资源支持软件更进一步发展；坏消息是 IBM 公司很大，可能很难与管理层讨价还价获得更多的优惠。另外，在你读到本书时可能学生版本已经没有了，但是你可以通过书店或者直接向 IBM 咨询。

需要更多的信息，请查阅官方网站。

价格：完整版是 2 249 美元，用于 Windows 系统的学术版本、学生版本是 70 美元，Mac 版本价格也大致相同。注意：学生版本是完整的基本软件包的修订版。

SYSTAT

生物学和心理学的研究者更常用 SYSTAT 13，而社会科学和行为科学的研究者喜欢 SPSS（虽然 SYSTAT 人员在最新发布的 13 版本中努力吸引社会和行为科学的研究者，增加了基本统计和方差分析）。这个软件只用于 Windows 系统（Mac 版本正在开发中）。这个软件支持很强的命令语言，因此分析可以适应使用者的需要。最新的版本支持新的界面和高度可定制的菜单。初学者可以使用这个软件，但是它更适合于高年级学生或专业人员。

需要更多的信息，请查阅官方网站。

价格：完整版本是 1 500 美元，你也可以下载 MYSTAT 版本，这是 SYSTAT 的免费版本，可以满足你的大多数需要。

STATISTIX for Windows

STATISTIX 9（只用于 Windows 系统）提供的主菜单驱动界面，使得这个软件特别容易学习和使用。几乎和这里提到的每个软件一样功能强大；公司还提供免费技术支持。不



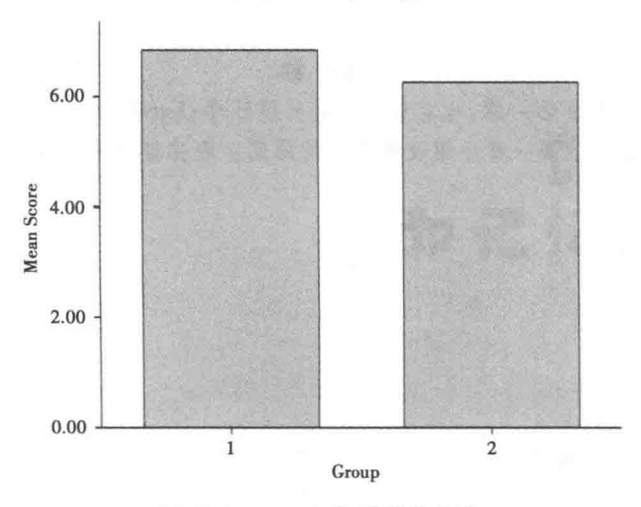


图 19.2 SPSS 运行的简单统计

但如此,还提供纸质的 400 页的使用手册——你想到这个没? 而且当你打电话寻求技术支持时,你可以和实际参与程序的设计人员谈话,他们知道你在说什么(我的问题在 10 秒内得到解答)。这个软件各部分都非常好。需要更多的信息,请查阅官方网站。

价格:学术版本是 395 美元。

EcStatic

在佛蒙特州的 Someware 公司工作人员的目标是"以合理的价格提供智能化、便捷型的统计和绘图软件"。其实他们做得更多。EsStatic 很能赚钱。EsStatic 是可以运行的软件中最便宜的,而且可以肯定的是相对于这里描述的大型软件支付的费用而言,这个软件物超所值。如果你认为这个软件遗漏了什么,看看下面列出的该软件可以分析的清单:

- 方差分析
- 数值转换
- 交互表和卡方
- 非参数统计
- 散点图
- 变量转换
- 分裂点
- 相关
- 频数分布和直方图
- 回归
- 汇总统计
- t 检验

下载试用版本体验一下吧!

需要更多的信息,请查阅官方网站。

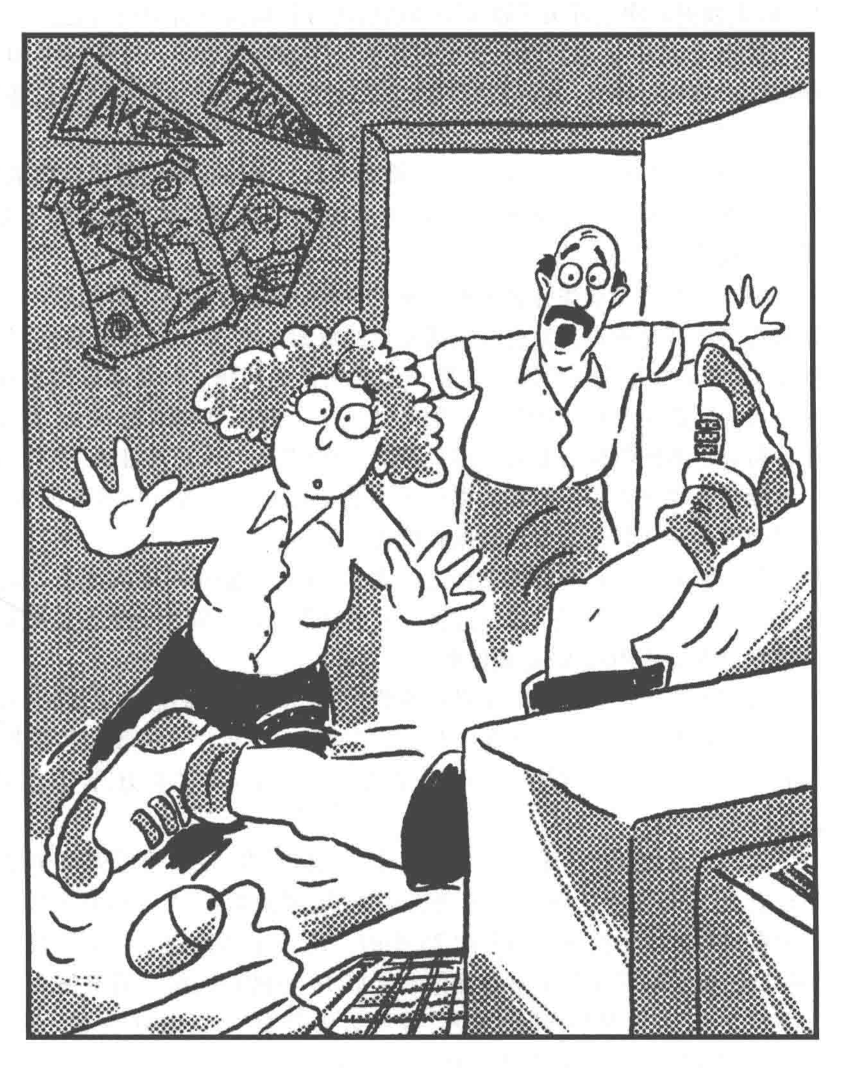
价格:99.95 美元就可以下载,还可以得到 10 美元或更多的折扣(69.95 美元)——告诉你的指导老师。

小结

这是第Ⅳ部分的最后一章，也应该是《爱上统计学：Excel》的最后部分。但是我认为你还需要继续往下读。下一章介绍收集数据的原则。祝你接下来的学习愉快！

第Ⅴ部分

你得了解和记忆的内容



“网络逮住他了……重启！重新启动！”

20 收集数据的10个原则

本书你学习的是如何分析数据,你还应该了解的是如何收集数据。数据收集过程很长也很严格,即使这个过程只是向学生、家长、病人或选举人等不同群体发放单页的简单问卷。数据收集过程也是你的研究项目中最耗时的部分。但是按照大多数研究者的经验,收集数据时期也可用于思考即将进行的分析以及分析中的问题。

这里给出10个原则确保你依据数据的用途收集数据。与最初的10个原则不同,这些原则不应该死记硬背(这些原则肯定会发生变化),但是如果你遵循这些原则,就可以避免很多失误。

原则1.在你思考研究问题之初,就要开始思考回答问题需要收集的数据类型。访谈?问卷调查?纸和笔?阅读你感兴趣的相关领域的期刊了解过去其他人如何收集数据并思考他们收集数据的目的。

原则2.在思考收集什么类型数据的同时,要思考从何处得到数据。如果是使用图书馆获得历史数据或者其他已经收集完全的数据文件,如人口调查数据(从美国人口调查局获得,或者上网在线获得),你可能不会遇到什么问题。但是如果你想评价新出生的孩子和父母之间的相互影响如何?教师对加入工会的态度?刚过50岁的人是否觉得自己老了?所有这些问题都需要人们提供答案,找到调查的人是很难的。要尽早开始。

原则3.保证你用于收集数据的表格简单容易。以飞行员的数据集进行练习,这样你可以保证从最初的记分表到数据收集表都很容易。

原则4.始终记得对数据文件进行备份,并且存储在不同的地方。记住有且只有两类人:一类已经丢失了数据,另一类将会丢失数据。在不同的位置备份数据收集表。如果你的数据以电子文档录入,也确保要进行备份。

原则5.不要依赖他人收集或转换数据,除非你对他们进行了培训,而且确信他们像你一样理解数据收集过程。让他人帮助你很好,而且在很长的收集数据期间对保持士气十分有帮助。但是,除非帮助你的人毋庸置疑地具备这样的能力,否则你很容易破坏你所有的工作和计划。

原则6.计划何时何地收集数据的详细的日程表。如果你需要参观3个学校而且每个学校有50个学生要分别进行10分钟的测试,那就需要25个小时进行测试。这并不意味着你在日程表中为这项活动分配25小时。从一个学校到另一个学校的时间呢?如果轮到测试的孩子刚好在卫生间而且你必须等10分钟直到他回到教室呢?或者你访问的这天刚好牛仔鲍伯是特别的客人,等等。为任何可能做好准备,要在日程表中分配25%~50%的多余时间应对不可预测的事件的发生。

原则7.只要有可能就为你的项目培育可能的数据来源。你已经对你的学科知识有了一定的了解,那么就可能知道谁和你需要的类型的人一起工作或者谁可能帮助你获得这

些样本。如果你是在大学社区，那么就可能有好几百人竞争你需要的相同的调查样本。如果不想竞争，为什么不尝试在学校之外（大约 30 分钟的距离）的社区、社会群体、市民组织或医院调查？在这些地方不需太多竞争就能够得到样本。

原则 8.尽力追踪遗漏了的测试或者访谈对象。把他们找回来并重新安排日程。一旦你习惯了跳过可能的参与者，那么样本规模将很容易缩小。而且你永远不能觉察——遗漏的人可能是由于与你的研究相关的原因而遗漏的，这意味着你最后的样本从性质上不同于你研究开始时设定的样本。

原则 9.永远不要销毁你的原始数据，如测试手册、访谈笔记等。其他的研究者可能想使用相同的数据库，或者你想要回到原始数据获得更多的信息。

原则 10.遵循前 9 个原则。这可不是开玩笑！

附录A　你应当掌握的Excel技能

如果你还不太会用Excel,这一章刚好是开始。我们列出了50个非常重要、非常有用的技能任务,这是你在开始就应该掌握的,这会让使用Excel进行数据分析变得非常容易。

我的建议是打开一个新的或者练习专用的工作表,这样就不会造成什么破坏,也可以输入新的数据。接着就是练习、练习、再练习。这是你学习Excel得到的最有用的提示:尽力尝试,不要害怕。你不会损坏任何东西,而且你只是学习,并获得乐趣。

Excel的新版本(包括2010版)的最好的设计之一是快捷访问工具栏。

你始终可以在Excel自定义工具栏,但是现在你可以在Excel功能区(也就是屏幕最上方的各种选项集)的上方建立某种汇总工具栏。你可以按照核心任务自定义工具栏,这样使得使用Excel更快且更容易,这就是我们要开始的地方。

按照下面的步骤在快捷工具栏增加一个图标。

1.点击[文件]选项,选择[选项],接着点击[自定义功能区]。

2.选中你想在屏幕左侧使用的命令,选中你想让其在屏幕的右侧出现的选项或者菜单。

3.点击[添加]按钮将命令添加到右侧的清单。

4.点击[确定],命令按钮就会出现在快捷访问工具栏。不要添加太多的选项或者菜单,只选择你想使用的命令。图A.1所示是快捷访问工具栏的样例,添加了新建、保存、撤销、恢复和打开(New, Save,Undo, Redo, Open)命令。非常好用。

图A.1　快捷访问工具栏样例:自定义工具栏节约时间

基础部分

1.建立新的工作表,点击[文件]选项,接着点击[新建]命令,或者按快捷键Ctrl+N。

2.打开已存在的工作表,点击[文件]选项,接着点击[打开]命令,或者使用快捷键Ctrl+O。

3.保存工作薄,点击[文件]选项,接着点击[保存]命令,或者使用快捷键Ctrl+S。

4.以不同的名称保存工作表,点击[文件]选项,接着点击[另存为]命令。

5.需要帮助? 按F键,或者在电子表的右上方点击问号符号。

6.搞错了? 使用快捷键Ctrl+Z返回你让Excel执行的上一个操作,可以用在数据录入或者大型的计算方面。

在一个工作表输入数据

7.在一个工作表输入数据,点击单元格,键入数据,然后按回车键。

8.输入序列数据,先输入序列数据的第一个数值,接着输入第二个数值,然后选中这两个数值,然后通过鼠标*将最后一个数值拖拽到最后一个单元格。(Excel 就自动生成一组序列数值,如输入 1,2,拖动鼠标可自动生成 1,2,3,4,……)

查找和替换数据

9.在工作表查找数据,点击[开始]选项,点击(编辑栏下)的[查找和选择],然后点击[查找]命令,或者使用快捷键 Ctrl+F,输入你想查找的数据,然后点击查找下一条。

10.查找和替换数据,点击(编辑栏下)的[查找和选择],点击[替换],或者使用快捷键Ctrl+H,输入你想查找的数据和你想替换成的数据,接着点击查找替换下一条。

编辑工作表

11.如果想选择一个单元格,就点击那个单元格。如果要选择整个工作表,点击 A 列和第一行之间的空格。

12.编辑单元格,点击某个单元格,按 F2 键,在这个单元格或者公式栏进行编辑,或者双击某个单元格,然后按照你的要求做出改变。

13.选择一行或者一列,点击行(数字)或者列(字母)的抬头。

14.插入一行或者一列,右击列字母或者行数,接着点击插入。你也可以使用开始选项中单元格分类选项下的插入和删除按钮。

15.复制单元格或者一定范围单元格的数据,选中要复制的数据,点击右键选择复制,点击你想放置复制数据的单元格,然后点击右键,选择粘贴。你也可以使用开始选项中的剪贴板,使用单个的命令完成相同的任务。

16. 复制单元格或者一定范围单元格的数据,选中要复制的数据,将鼠标的指示光标放置在选中的单元格边界,然后按 Crtl 键,光标会以一个手型出现,接着拖拽数据到新的位置。

17.替换数据,点击这个单元格,输入新的数据,然后按回车键。

18.剪切单元格或者一定范围单元格的数据,选中要剪切的数据,点击右键,选中剪切,或者按删除键。

19.设置数据的格式,选择单元格,点击右键选择设置单元格格式,点击数字选项。选择数字格式,然后点击确定。

---

* 光标置于选中的两个单元格的右下角,呈现为黑色十字形。——译者注

设置工作表的样式

20.在工作表设置单元格的行和列的适宜数据，选中整个工作表，点击开始选项，在单元格分项中点击自动调整列宽和自动调整行高。

21.改变数据的格式，选中你想改变的单元格，点击右键选中设置单元格格式，然后做出改变。

22.改变默认字体，点击文件选项，点击其他命令分项，选择字体和字号。

23.增加单元格边框，选中你想进行改变的单元格，在字体分项中的边框选项下拉菜单中点击你想要的边框类型。

24.给一个单元格增加阴影，点击开始选项，然后在字体分项中，点击小油漆桶的下拉菜单，选择你想用作阴影的颜色。

25.改变工作表的边界，点击[文件]选项→[打印]→[页面设置]，然后定义边界。

26.增加工作表的页眉和页脚，点击插入选项，选择文本分项，点击页眉和页脚，然后输入定义的页眉和或页脚。然后点击页眉或页脚之外的其他任何区域。

27.显示或隐藏网格线，点击视图选项，在显示/隐藏分项中点击网格线复选框。

28.加粗字体，选中要加粗的部分，在开始选项的字体分项中点击加粗按钮，或者使用快捷键 Ctrl+B。

29.倾斜字体，选中要倾斜的部分，在开始选项的字体分项中点击倾斜按钮，或者使用快捷键 Ctrl+I。

30.改变单元格对齐方式，在开始选项的对齐方式分享中选择左对齐、居中或者右对齐。

31.改变数字的小数点位数，在开始选项的数字分项中选择增加小数点位数或者减少小数点位数。

32.删除数据中间的空行，分类或者筛选数据，最后删除空行。

33. 在任何单元格你想换行的位置点击[自动换行]，就可以建立新的一行。

关于单元格和数值

34.命名单元格区域，选择你想命名的单元格，点击右键，点击命名单元格区域，输入名称，然后点击 OK。

35.建立公式，点击你想让公式结果返回的单元格，输入等号(=)，输入公式，然后按回车键。

36.输入函数，点击你想让函数结果返回的单元格，点击函数选项，点击你想使用的函数所在的函数类型，选择这个函数。

37.分类和排序数据，选择你想进行分类和排序的单元格，点击数据选项，选择升序或者降序分类和排序。

38.隐藏行或者列，选择你想隐藏的行或者列，点击开始选项，在单元格分项中选择隐藏行或者列。

39.取消隐藏行或者列,选择距离隐藏的行或者列最近的行或者列,点击开始选项,在单元格分项中选择取消隐藏行或者列。

40.数字进行平方,使用^符号;例如输入 = 2^2。求平方根,使用1/2作为幂,例如,输入2^(1/2)。

使用分析工具库

41.使用分析工具库,点击数据选项,点击分析分项中选择数据分析,双击你想进行的分析。

建立和使用图表

42.建立一个图表,选中数据,包括列和行的标签;点击插入选项,在图表分项中点击你想建立的图表类型。

43.调整图表的大小,水平、垂直或者对角拖动图表边框调整图表大小。

44.改变图表的标题、标签或者其他关于图表的信息,点击图表工具下的布局选项,然后点击你想改变的信息。

45.只打印图表,在图表上点击选中图表,点击文件按钮,选择打印。

46.改变图表要素的格式,点击要素,如最大或者最小刻度;按照你的要求进行改变;然后点击确定。

47.复制图表到其他应用软件,按快捷键Ctrl+A选中图表,Ctrl+C复制图表,在新的应用软件中,按快捷键Ctrl+V。

打印工作表

48.打印工作表,点击[文件]选项,然后点击[打印]。

49.预览工作表,点击[文件]选项,点击[打印预览],你可以看到预览工作表。

50.打印工作表中不连接在一起的单元格,按住Ctrl键的同时选择你要打印的单元格。

附录B 统计值表

表 B.1：正态曲线下的面积

如何使用这个表：

1.依据样本的原始数值和均值计算 z 值。

2.依据 z 值确定正态曲线下面积的百分比或者确定均值和计算的 z 值之间面积的百分比。

表 B.2：拒绝零假设需要的 t 值

如何使用这个表：

1.计算检验统计量 t 值。

2.比较实际值 t 值和这个表中的临界值。确定你正确地计算了自由度，而且选择了合适的显著水平。

3.如果实际值大于临界值或这个表中的值，零假设（均值相等）就不是观察到的任何差异的最有力解释。

4.如果实际值小于临界值或这个表中的值，零假设就是观察到的任何差异的最有力解释。

表 B.3：方差分析或者 F 检验的临界值

如何使用这个表：

1.计算 F 值。

2.计算分子的自由度（$k-1$）和分母的自由度（$n-k$）。

3.依据分子自由度和分母自由度交叉的位置确定临界值。临界值就是行和列交叉位置的值。

4.如果实际值大于临界值或这个表中的值，零假设（均值彼此相等）就不是观察到的任何差异的最有力解释。

5.如果实际值小于临界值或这个表中的值，零假设就是观察到的任何差异的最有力解释。

表 B.4：拒绝零假设需要的相关系数值

如何使用这个表：

1.计算相关系数值。

2.比较相关系数值和这个表中的临界值。

3.如果实际值大于临界值或这个表中的值,零假设(相关系数等于 0)就不是观察到的任何差异的最有力解释。

4.如果实际值小于临界值或这个表中的值,零假设就是观察到的任何差异的最有力解释。

表 B.5:卡方检验的临界值

如何使用这个表:

1.计算 χ^2 值。

2.计算行的自由度($R-1$)和列的自由度($C-1$)。如果是一维表,就只有列的自由度。

3.依据标题为(df)列的自由度和合适的显著水平所在的列确定对应的临界值。

4.如果实际值大于临界值或这个表中的值,零假设(频数彼此相等)就不是观察到的任何差异的最有力解释。

5.如果实际值小于临界值或这个表中的值,零假设就是观察到的任何差异的最有力解释。

表 B.1　正态曲线下的面积

z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积
0.00	0.00	0.50	19.15	1.00	34.13	1.50	43.32	2.00	47.72	2.50	49.38	3.00	49.87	3.50	49.98
0.01	0.40	0.51	19.50	1.01	34.38	1.51	43.45	2.01	47.78	2.51	49.40	3.01	49.87	3.51	49.98
0.02	0.50	0.52	19.85	1.02	34.61	1.52	43.57	2.02	47.83	2.52	49.41	3.02	49.87	3.52	49.98
0.03	1.20	0.53	20.19	1.03	34.85	1.53	43.70	2.03	47.88	2.53	49.43	3.03	49.88	3.53	49.98
0.04	1.60	0.54	20.54	1.04	35.08	1.54	43.82	2.04	47.93	2.54	49.45	3.04	49.88	3.54	49.98
0.05	1.99	0.55	20.88	1.05	35.31	1.55	43.94	2.05	47.98	2.55	49.46	3.05	49.89	3.55	49.98
0.06	2.39	0.56	21.23	1.06	35.54	1.56	44.06	2.06	48.03	2.56	49.48	3.06	49.89	3.56	49.98
0.07	2.79	0.57	21.57	1.07	35.77	1.57	44.18	2.07	48.08	2.57	49.49	3.07	49.89	3.57	49.98
0.08	3.19	0.58	21.90	1.08	35.99	1.58	44.29	2.08	48.12	2.58	49.51	3.08	49.90	3.58	49.98
0.09	3.59	0.59	22.24	1.09	36.21	1.59	44.41	2.09	48.17	2.59	49.52	3.09	49.90	3.59	49.98
0.10	3.98	0.60	22.57	1.10	36.43	1.60	44.52	2.10	48.21	2.60	49.53	3.10	49.90	3.60	49.98
0.11	4.38	0.61	22.91	1.11	36.65	1.61	44.63	2.11	48.26	2.61	49.55	3.11	49.91	3.61	49.98
0.12	4.78	0.62	23.24	1.12	36.86	1.62	44.74	2.12	48.30	2.62	49.56	3.12	49.91	3.62	49.98
0.13	5.17	0.63	23.57	1.13	37.08	1.63	44.84	2.13	48.34	2.63	49.57	3.13	49.91	3.63	49.98
0.14	5.57	0.64	23.89	1.14	37.29	1.64	44.95	2.14	48.38	2.64	49.59	3.14	49.92	3.64	49.98
0.15	5.96	0.65	24.22	1.15	37.49	1.65	45.05	2.15	48.42	2.65	49.60	3.15	49.92	3.65	49.98
0.16	6.36	0.66	24.54	1.16	37.70	1.66	45.15	2.16	48.46	2.66	49.61	3.16	49.92	3.66	49.98
0.17	6.75	0.67	24.86	1.17	37.90	1.67	45.25	2.17	48.50	2.67	49.62	3.17	49.92	3.67	49.98
0.18	7.14	0.68	25.18	1.18	38.10	1.68	45.35	2.18	48.54	2.68	49.63	3.18	49.93	3.68	49.98
0.19	7.53	0.69	25.49	1.19	38.30	1.69	45.45	2.19	48.57	2.69	49.64	3.19	49.93	3.69	49.98
0.20	7.93	0.70	25.80	1.20	38.49	1.70	45.54	2.20	48.61	2.70	49.65	3.20	49.93	3.70	49.99
0.21	8.32	0.71	26.11	1.21	38.69	1.71	45.64	2.21	48.64	2.71	49.66	3.21	49.93	3.71	49.99
0.22	8.71	0.72	26.42	1.22	38.88	1.72	45.73	2.22	48.68	2.72	49.67	3.22	49.94	3.72	49.99
0.23	9.10	0.73	26.73	1.23	39.07	1.73	45.82	2.23	48.71	2.73	49.68	3.23	49.94	3.73	49.99
0.24	9.48	0.74	27.04	1.24	39.25	1.74	45.91	2.24	48.75	2.74	49.69	3.24	49.94	3.74	49.99
0.25	9.99	0.75	27.34	1.25	39.44	1.75	45.99	2.25	48.78	2.75	49.70	3.25	49.94	3.75	49.99
0.26	10.26	0.76	27.64	1.26	39.62	1.76	46.08	2.26	48.81	2.76	49.71	3.26	49.94	3.76	49.99
0.27	10.64	0.77	27.94	1.27	39.80	1.77	46.16	2.27	48.84	2.77	49.72	3.27	49.94	3.77	49.99
0.28	11.03	0.78	28.23	1.28	39.97	1.78	46.25	2.28	48.87	2.78	49.73	3.28	49.94	3.78	49.99
0.29	11.41	0.79	28.52	1.29	40.15	1.79	46.33	2.29	48.90	2.79	49.74	3.29	49.94	3.79	49.99
0.30	11.79	0.80	28.81	1.30	40.32	1.80	46.41	2.30	48.93	2.80	49.74	3.30	49.95	3.80	49.99
0.31	12.17	0.81	29.10	1.31	40.49	1.81	46.49	2.31	48.96	2.81	49.75	3.31	49.95	3.81	49.99

续表

z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积	z 值	均值和 z 值之间的面积
0.32	12.55	0.82	29.39	1.32	40.66	1.82	46.56	2.32	48.98	2.82	49.76	3.32	49.95	3.82	49.99
0.33	12.93	0.83	29.67	1.33	40.82	1.83	46.64	2.33	49.01	2.83	49.77	3.33	49.95	3.83	49.99
0.34	13.31	0.84	29.95	1.34	40.99	1.84	46.71	2.34	49.04	2.84	49.77	3.34	49.95	3.84	49.99
0.35	13.68	0.85	30.23	1.35	41.15	1.85	46.78	2.35	49.06	2.85	49.78	3.35	49.96	3.85	49.99
0.36	14.06	0.86	30.51	1.36	41.31	1.86	46.86	2.36	49.09	2.86	49.79	3.36	49.96	3.86	49.99
0.37	14.43	0.87	30.78	1.37	41.47	1.87	46.93	2.37	49.11	2.87	49.79	3.37	49.96	3.87	49.99
0.38	14.80	0.88	31.06	1.38	41.62	1.88	46.99	2.38	49.13	2.88	49.80	3.38	49.96	3.88	49.99
0.39	15.17	0.89	31.33	1.39	41.77	1.89	47.06	2.39	49.16	2.89	49.81	3.39	49.96	3.89	49.99
0.40	15.54	0.90	31.59	1.40	41.92	1.90	47.13	2.40	49.18	2.90	49.81	3.40	49.97	3.90	49.99
0.41	15.91	0.91	31.86	1.41	42.07	1.91	47.19	2.41	49.20	2.91	49.82	3.41	49.97	3.91	49.99
0.42	16.28	0.92	32.12	1.42	42.22	1.92	47.26	2.42	49.22	2.92	49.82	3.42	49.97	3.92	49.99
0.43	16.64	0.93	32.38	1.43	42.36	1.93	47.32	2.43	49.25	2.93	49.83	3.43	49.97	3.93	49.99
0.44	17.00	0.94	32.64	1.44	42.51	1.94	47.38	2.44	49.27	2.94	49.84	3.44	49.97	3.94	49.99
0.45	17.36	0.95	32.89	1.45	42.65	1.95	47.44	2.45	49.29	2.95	49.84	3.45	49.98	3.95	49.99
0.46	17.72	0.96	33.15	1.46	42.79	1.96	47.50	2.46	49.31	2.96	49.85	3.46	49.98	3.96	49.99
0.47	18.08	0.97	33.40	1.47	42.92	1.97	47.56	2.47	49.32	2.97	49.85	3.47	49.98	3.97	49.99
0.48	18.44	0.98	33.65	1.48	43.06	1.98	47.61	2.48	49.34	2.98	49.86	3.48	49.98	3.98	49.99
0.49	18.79	0.99	33.89	1.49	43.19	1.99	47.67	2.49	49.36	2.99	49.86	3.49	49.98	3.99	49.99

表 B.2 拒绝零假设需要的 t 值

自由度	单侧检验			自由度	双侧检验		
	0.10	**0.05**	**0.01**		**0.10**	**0.05**	**0.01**
1	3.078	6.314	31.821	1	6.314	12.706	63.657
2	1.886	2.92	6.965	2	2.92	4.303	9.925
3	1.638	2.353	4.541	3	2.353	3.182	5.841
4	1.533	2.132	3.747	4	2.132	2.776	4.604
5	1.476	2.015	3.365	5	2.015	2.571	4.032
6	1.44	1.943	3.143	6	1.943	2.447	3.708
7	1.415	1.895	2.998	7	1.895	2.365	3.5
8	1.397	1.86	2.897	8	1.86	2.306	3.356
9	1.383	1.833	2.822	9	1.833	2.262	3.25
10	1.372	1.813	2.764	10	1.813	2.228	3.17
11	1.364	1.796	2.718	11	1.796	2.201	3.106
12	1.356	1.783	2.681	12	1.783	2.179	3.055
13	1.35	1.771	2.651	13	1.771	2.161	3.013
14	1.345	1.762	2.625	14	1.762	2.145	2.977
15	1.341	1.753	2.603	15	1.753	2.132	2.947
16	1.337	1.746	2.584	16	1.746	2.12	2.921
17	1.334	1.74	2.567	17	1.74	2.11	2.898
18	1.331	1.734	2.553	18	1.734	2.101	2.879
19	1.328	1.729	2.54	19	1.729	2.093	2.861
20	1.326	1.725	2.528	20	1.725	2.086	2.846
21	1.323	1.721	2.518	21	1.721	2.08	2.832
22	1.321	1.717	2.509	22	1.717	2.074	2.819
23	1.32	1.714	2.5	23	1.714	2.069	2.808
24	1.318	1.711	2.492	24	1.711	2.064	2.797
25	1.317	1.708	2.485	25	1.708	2.06	2.788
26	1.315	1.706	2.479	26	1.706	2.056	2.779
27	1.314	1.704	2.473	27	1.704	2.052	2.771
28	1.313	1.701	2.467	28	1.701	2.049	2.764
29	1.312	1.699	2.462	29	1.699	2.045	2.757
30	1.311	1.698	2.458	30	1.698	2.043	2.75
35	1.306	1.69	2.438	35	1.69	2.03	2.724
40	1.303	1.684	2.424	40	1.684	2.021	2.705
45	1.301	1.68	2.412	45	1.68	2.014	2.69
50	1.299	1.676	2.404	50	1.676	2.009	2.678
55	1.297	1.673	2.396	55	1.673	2.004	2.668
60	1.296	1.671	2.39	60	1.671	2.001	2.661
65	1.295	1.669	2.385	65	1.669	1.997	2.654
70	1.294	1.667	2.381	70	1.667	1.995	2.648
75	1.293	1.666	2.377	75	1.666	1.992	2.643
80	1.292	1.664	2.374	80	1.664	1.99	2.639
85	1.292	1.663	2.371	85	1.663	1.989	2.635
90	1.291	1.662	2.369	90	1.662	1.987	2.632
95	1.291	1.661	2.366	95	1.661	1.986	2.629
100	1.29	1.66	2.364	100	1.66	1.984	2.626
无穷大	1.282	1.645	2.327	无穷大	1.645	1.96	2.576

表 B.3　方差分析或 F 检验的临界值

分母自由度	第一类错误	分子自由度					
		1	**2**	**3**	**4**	**5**	**6**
1	.01	4052.00	4999.00	5403.00	5625.00	5764.00	5859.00
	.05	162.00	200.00	216.00	225.00	230.00	234.00
	.10	39.90	49.50	53.60	55.80	57.20	58.20
2	.01	98.50	99.00	99.17	99.25	99.30	99.33
	.05	18.51	19.00	19.17	19.25	19.30	19.33
	.10	8.53	9.00	9.16	9.24	9.29	9.33
3	.01	34.12	30.82	29.46	28.71	28.24	27.91
	.05	10.13	9.55	9.28	9.12	9.01	8.94
	.10	5.54	5.46	5.39	5.34	5.31	5.28
4	.01	21.20	18.00	16.70	15.98	15.52	15.21
	.05	7.71	6.95	6.59	6.39	6.26	6.16
	.10	.55	4.33	4.19	4.11	4.05	4.01
5	.01	16.26	13.27	12.06	11.39	10.97	10.67
	.05	6.61	5.79	5.41	5.19	5.05	4.95
	.10	4.06	3.78	3.62	3.52	3.45	3.41
6	.01	13.75	10.93	9.78	9.15	8.75	8.47
	.05	5.99	5.14	4.76	4.53	4.39	4.28
	.10	3.78	3.46	3.29	3.18	3.11	3.06
7	.01	12.25	9.55	8.45	7.85	7.46	7.19
	.05	5.59	4.74	4.35	4.12	3.97	3.87
	.10	3.59	3.26	3.08	2.96	2.88	2.83
8	.01	11.26	8.65	7.59	7.01	6.63	6.37
	.05	5.32	4.46	4.07	3.84	3.69	3.58
	.10	3.46	3.11	2.92	2.81	2.73	2.67
9	.01	10.56	8.02	6.99	6.42	6.06	5.80
	.05	5.12	4.26	3.86	3.63	3.48	3.37
	.10	3.36	3.01	2.81	2.69	2.61	2.55
10	.01	10.05	7.56	6.55	6.00	5.64	5.39
	.05	4.97	4.10	3.71	3.48	3.33	3.22
	.10	3.29	2.93	2.73	2.61	2.52	2.46
11	.01	9.65	7.21	6.22	5.67	5.32	5.07
	.05	4.85	3.98	3.59	3.36	3.20	3.10
	.10	3.23	2.86	2.66	2.54	2.45	2.39
12	.01	9.33	6.93	5.95	5.41	5.07	4.82
	.05	4.75	3.89	3.49	3.26	3.11	3.00
	.10	3.18	2.81	2.61	2.48	2.40	2.33

注:此表因软件原因,小数点前的数字 0 未显示。

续表

分母自由度	第一类错误	分子自由度					
		1	**2**	**3**	**4**	**5**	**6**
13	.01	9.07	6.70	5.74	5.21	4.86	4.62
	.05	4.67	3.81	3.41	3.18	3.03	2.92
	.10	3.14	2.76	2.56	2.43	2.35	2.28
14	.01	8.86	6.52	5.56	5.04	4.70	4.46
	.05	4.60	3.74	3.34	3.11	2.96	2.85
	.10	3.10	2.73	2.52	2.40	2.31	2.24
15	.01	8.68	6.36	5.42	4.89	4.56	4.32
	.05	4.54	3.68	3.29	3.06	2.90	2.79
	.10	3.07	2.70	2.49	2.36	2.27	2.21
16	.01	8.53	6.23	5.29	4.77	4.44	4.20
	.05	4.49	3.63	3.24	3.01	2.85	2.74
	.10	3.05	2.67	2.46	2.33	2.24	2.18
17	.01	8.40	6.11	5.19	4.67	4.34	4.10
	.05	4.45	3.59	3.20	2.97	2.81	2.70
	.10	3.03	2.65	2.44	2.31	2.22	2.15
18	.01	8.29	6.01	5.09	4.58	4.25	4.02
	.05	4.41	3.56	3.16	2.93	2.77	2.66
	.10	3.01	2.62	2.42	2.29	2.20	2.13
19	.01	8.19	5.93	5.01	4.50	4.17	3.94
	.05	4.38	3.52	3.13	2.90	2.74	2.63
	.10	2.99	2.61	2.40	2.27	2.18	2.11
20	.01	8.10	5.85	4.94	4.43	4.10	3.87
	.05	4.35	3.49	3.10	2.87	2.71	2.60
	.10	2.98	2.59	2.38	2.25	2.16	2.09
21	.01	8.02	5.78	4.88	4.37	4.04	3.81
	.05	4.33	3.47	3.07	2.84	2.69	2.57
	.10	2.96	2.58	2.37	2.23	2.14	2.08
22	.01	7.95	5.72	4.82	4.31	3.99	3.76
	.05	4.30	3.44	3.05	2.82	2.66	2.55
	.10	2.95	2.56	2.35	2.22	2.13	2.06
23	.01	7.88	5.66	4.77	4.26	3.94	3.71
	.05	4.28	3.42	3.03	2.80	2.64	2.53
	.10	2.94	2.55	2.34	2.21	2.12	2.05
24	.01	7.82	5.61	4.72	4.22	3.90	3.67
	.05	4.26	3.40	3.01	2.78	2.62	2.51
	.10	2.93	2.54	2.33	2.20	2.10	2.04

续表

分母自由度	第一类错误	分子自由度					
		1	**2**	**3**	**4**	**5**	**6**
25	.01	7.77	5.57	4.68	4.18	3.86	3.63
	.05	4.24	3.39	2.99	2.76	2.60	2.49
	.10	2.92	2.53	2.32	2.19	2.09	2.03
26	.01	7.72	5.53	4.64	4.14	3.82	3.59
	.05	4.23	3.37	2.98	2.74	2.59	2.48
	.10	2.91	2.52	2.31	2.18	2.08	2.01
27	.01	7.68	5.49	4.60	4.11	3.79	3.56
	.05	4.21	3.36	2.96	2.73	2.57	2.46
	.10	2.90	2.51	2.30	2.17	2.07	2.01
28	.01	7.64	5.45	4.57	4.08	3.75	3.53
	.05	4.20	3.34	2.95	2.72	2.56	2.45
	.10	2.89	2.50	2.29	2.16	2.07	2.00
29	.01	7.60	5.42	4.54	4.05	3.73	3.50
	.05	4.18	3.33	2.94	2.70	2.55	2.43
	.10	2.89	2.50	2.28	2.15	2.06	1.99
30	.01	7.56	5.39	4.51	4.02	3.70	3.47
	.05	4.17	3.32	2.92	2.69	2.53	2.42
	.10	2.88	2.49	2.28	2.14	2.05	1.98
35	.01	7.42	5.27	4.40	3.91	3.59	3.37
	.05	4.12	3.27	2.88	2.64	2.49	2.37
	.10	2.86	2.46	2.25	2.14	2.02	1.95
40	.01	7.32	5.18	4.31	3.91	3.51	3.29
	.05	4.09	3.23	2.84	2.64	2.45	2.34
	.10	2.84	2.44	2.23	2.11	2.00	1.93
45	.01	7.23	5.11	4.25	3.83	3.46	3.23
	.05	4.06	3.21	2.81	2.61	2.42	2.31
	.10	2.82	2.43	2.21	2.09	1.98	1.91
50	.01	7.17	5.06	4.20	3.77	3.41	3.19
	.05	4.04	3.18	2.79	2.58	2.40	2.29
	.10	2.81	2.41	2.20	2.08	1.97	1.90
55	.01	7.12	5.01	4.16	3.72	3.37	3.15
	.05	4.02	3.17	2.77	2.56	2.38	2.27
	.10	2.80	2.40	2.19	2.06	1.96	1.89
60	.01	7.08	4.98	4.13	3.68	3.34	3.12
	.05	4.00	3.15	2.76	2.54	2.37	2.26
	.10	2.79	2.39	2.18	2.05	1.95	1.88

续表

分母自由度	第一类错误	分子自由度					
		1	**2**	**3**	**4**	**5**	**6**
65	.01	7.04	4.95	4.10	3.65	3.31	3.09
	.05	3.99	3.14	2.75	2.53	2.36	2.24
	.10	2.79	2.39	2.17	2.04	1.94	1.87
70	.01	7.01	4.92	4.08	3.62	3.29	3.07
	.05	3.98	3.13	2.74	2.51	2.35	2.23
	.10	2.78	2.38	2.16	2.03	1.93	1.86
75	.01	6.99	4.90	4.06	3.60	3.27	3.05
	.05	3.97	3.12	2.73	2.50	2.34	2.22
	.10	2.77	2.38	2.16	2.03	1.93	1.86
80	.01	3.96	4.88	4.04	3.56	3.26	3.04
	.05	6.96	3.11	2.72	2.49	2.33	2.22
	.10	2.77	2.37	2.15	2.02	1.92	1.85
85	.01	6.94	4.86	4.02	3.55	3.24	3.02
	.05	3.95	3.10	2.71	2.48	2.32	2.21
	.10	2.77	2.37	2.15	2.01	1.92	1.85
90	.01	6.93	4.85	4.02	3.54	3.23	3.01
	.05	3.95	3.10	2.71	2.47	2.32	2.20
	.10	2.76	2.36	2.15	2.01	1.91	1.84
95	.01	6.91	4.84	4.00	3.52	3.22	3.00
	.05	3.94	3.09	2.70	2.47	2.31	2.20
	.10	2.76	2.36	2.14	2.01	1.91	1.84
100	.01	6.90	4.82	3.98	3.51	3.21	2.99
	.05	3.94	3.09	2.70	2.46	2.31	2.19
	.10	2.76	2.36	2.14	2.00	1.91	1.83
无穷大	.01	6.64	4.61	3.78	3.32	3.02	2.80
	.05	3.84	3.00	2.61	2.37	2.22	2.10
	.10	2.71	2.30	2.08	1.95	1.85	1.78

表 B.4 拒绝零假设需要的相关系数值

自由度	单侧检验		自由度	双侧检验	
	.05	**.01**		**.05**	**.01**
1	.9877	.9995	1	.9969	.9999
2	.9000	.9800	2	.9500	.9900
3	.8054	.9343	3	.8783	.9587
4	.7293	.8822	4	.8114	.9172
5	.6694	.832	5	.7545	.8745
6	.6215	.7887	6	.7067	.8343
7	.5822	.7498	7	.6664	.7977
8	.5494	.7155	8	.6319	.7646
9	.5214	.6851	9	.6021	.7348
10	.4973	.6581	10	.5760	.7079
11	.4762	.6339	11	.5529	.6835
12	.4575	.6120	12	.5324	.6614
13	.4409	.5923	13	.5139	.6411
14	.4259	.5742	14	.4973	.6226
15	.412	.5577	15	.4821	.6055
16	.4000	.5425	16	.4683	.5897
17	.3887	.5285	17	.4555	.5751
18	.3783	.5155	18	.4438	.5614
19	.3687	.5034	19	.4329	.5487
20	.3598	.4921	20	.4227	.5368
25	.3233	.4451	25	.3809	.4869
30	.2960	.4093	30	.3494	.4487
35	.2746	.3810	35	.3246	.4182
40	.2573	.3578	40	.3044	.3932
45	.2428	.3384	45	.2875	.3721
50	.2306	.3218	50	.2732	.3541
60	.2108	.2948	60	.2500	.3248
70	.1954	.2737	70	.2319	.3017
80	.1829	.2565	80	.2172	.2830
90	.1726	.2422	90	.2050	.2673
100	.1638	.2301	100	.1946	.2540

注：此表因软件原因，小数点前的数字 0 未显示。

表 B.5 卡方检验的临界值

自由度	显著水平		
	.10	**.05**	**.01**
1	2.71	3.84	6.64
2	4.00	5.99	9.21
3	6.25	7.82	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	16.99	18.31	23.21
11	17.28	19.68	24.72
12	18.65	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.60	33.41
18	25.99	28.87	34.80
19	27.20	30.14	36.19
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.42	42.98
25	34.38	37.65	44.81
26	35.56	38.88	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59
30	40.26	43.77	50.89

附录C 练习题数据集*

第2章数据集1

	Score 1	Score 2	Score 3
	3	34	154
	7	54	167
	5	17	132
	4	26	145
	5	34	154
	6	25	145
	7	14	113
	8	24	156
	6	25	154
	5	23	123
Mean	5.6	27.6	144.3
Median	5.5	25.0	149.5
Mode	5.0	34.0	154.0

第2章数据集2

Day	Class 1	Class 2	Class 3
Monday	23	32	17
Tuesday	15	44	12
Wednesday	34	15	27
Thursday	32	44	19
Friday	33	36	23

第3章数据集1

Height	Weight	Height	Weight
53	156	57	154
46	131	68	166
54	123	65	153
44	142	66	140
56	156	54	143
76	171	66	156
87	143	51	173
65	135	58	143
45	138	49	161
44	114	48	131

* 练习题数据的 Excel 文件可扫码封底二维码获取。

第 3 章数据集 2

No Intervention	Intervention
18	17
16	15
17	21
15	24
16	12
14	16
22	14
21	25
7	19
24	24
23	21
19	25
12	9
20	9
23	13

第 4 章数据集 1

Comp Score	Comp Score	Comp Score	Comp Score
12	36	49	29
15	34	45	54
11	33	45	56
16	38	47	57
21	42	43	59
25	44	31	54
21	47	12	56
8	54	14	43
6	55	15	44
2	51	16	41
22	56	22	42
26	53	29	7
27	57		

第 5 章数据集 1

Correct	Attitude
17	94
13	73
12	59
15	80
16	93
14	85
16	66
16	79
18	77
19	91

第 5 章数据集 2

Years of Training	Successful Outcomes	Yearsof Training	Successful Outcomes
1	9	7	9
9	1	9	5
1	8	7	5
4	7	6	6
3	6	6	7
3	7	1	4

第 5 章数据集 3

Wash	Number	Infect

第 6 章数据集 1

Fall Results	Spring Results	Fall Results	Spring Results
21	7	3	30
38	13	16	26
15	35	34	43
34	45	50	20
5	19	14	22
32	47	14	25
24	34	3	50
3	1	4	17
17	12	42	32
32	41	28	46
33	3	40	10
15	20	40	48
21	39	12	11
8	46	5	23

第 8 章数据集 1

z Score
1.5
2.1
0.7
0.4
2
1.6
0.4
0.8
0.5
1.7

第 10 章数据集 1

61	0.826
78	
73	
92	
71	
83	
58	
78	
94	
57	
74	
76	
62	
91	
53	
60	
97	
55	
92	
60	
93	

100	
65	
53	
54	
66	
73	
89	
61	
57	
60	
51	
99	
77	
74	
58	
58	
71	
74	
92	
50	

73	
63	
59	
93	
94	
72	
89	
88	
66	
83	
71	
53	
87	
66	
87	
99	
57	
64	
57	
98	

51	
89	
100	
81	
61	
68	
92	
64	
93	
93	
76	
83	
70	
52	
86	
55	
70	
94	
91	
59	

73	
98	
81	
86	
61	
75	
86	
88	
93	
54	
51	
69	
63	
71	
92	
71	
97	
72	
92	
74	

第 10 章数据集 2

Swimming Time	Swimming Time	Swimming Time	Swimming Time
22	24	24	24
23	22	25	23
20	24	20	23
23	24	22	23
20	24	24	20
23	21	24	21
24	24	22	24
21	25	22	24
21	21	20	23
21	20	25	20
24	21	21	23
24	23	21	25
25	21	24	24
25	21	25	23
24	22	25	22
21	25	24	24
24	23	20	24
25	24	23	25
20	21	25	25
24	25	22	23
20	21	21	22
20	20	22	23
20	21	21	25
25	22	24	25
24	24	25	22
21	23	23	21
21	24	24	22
23	22	23	21
24	24	22	20
21	25	24	21
22	23	22	21
22	22	25	21
22	25	21	23
22	25	22	22
25	22	22	21
23	21	21	24
25	20	23	24
22	21	23	20
23	23	20	20
21	22	20	24
21	23	23	23
25	20	25	25
23	22	23	20
20	24	25	22
22	20	23	25
21	20	25	20
25	22	24	23
25	23	25	22
20	25	20	25
22	20	21	22
23	22	23	20
21	24	21	20

续表

Swimming Time	Swimming Time	Swimming Time	Swimming Time
22	25	21	20
22	25	22	21
22	22	23	22
20	20	24	21
23	23	23	21
21	21	25	21
24	25	21	21
23	22	25	22
23	20	21	21
24	20	20	23
22	25	24	20
23	24	24	21
21	20	24	25
23	21	20	21
23	20	25	23
20	20	21	20
24	23	25	25
25	22	23	20
24	20	25	20
22	21	24	24
22	25	20	20
21	23	21	23
20	20	20	20
20	21	21	24
25	20	23	24
21	20	20	21
21	20	21	25
23	25	21	20
20	25	21	21
21	22	23	25
22	21	24	20
20	21	21	23
25	21	20	20
20	22	23	21
23	20	25	24
21	21	20	23
21	20	21	25
24	22	20	22
21	25	24	25
22	23	22	20
22	24	21	22
20	25	25	21
22	21	21	23
23	22	21	25
21	25	20	23
23	22	22	24
23	22	23	20
22	20	24	24
23	23	23	20
21	21	23	21
20	23	22	20
20	25	25	24

续表

Swimming Time	Swimming Time	Swimming Time	Swimming Time
21	22	20	25
24	24	20	24
25	24	21	21
20	24	20	20
21	24	23	22
25	23	25	21
20	23	23	23
24	23	22	24
24	22	21	22
24	24	20	20
22	22	23	24
25	23	25	24
24	21	25	25
20	22	21	20
25	23	20	23
21	25	23	24
20	25	25	23
25	20	24	22
24	24	24	24
20	22	20	24
25	24	20	20
25	24	22	23
22	20	22	21
21	23	24	24
23	20	24	24
23	21	20	23
24	22	22	24
23	21	24	25
22	20	23	21
25	23	25	22
24	25	20	23
24	23	21	20
24	20	24	24
23	22	23	22
25	23	21	20
20	23	21	20
21	24	23	21
23	25	21	21
23	21	23	20
25	25	21	20
25	24	21	21
25	20	20	21
24	23	21	24
21	22	23	21
21	24	20	25
22	24	21	25
22	24	25	20
23	21	21	23
23	24	22	23
24	21	24	20
20	21	23	23
21	25	20	21

续表

Swimming Time	Swimming Time	Swimming Time	Swimming Time
24	23	23	24
21	22	21	20
25	20	22	25
20	21	25	25
24	25	22	22
24	25	23	20
21	22	21	25
20	20	22	24
25	25	20	21
20	24	21	21
21	21	23	23
20	22	23	21
20	22	22	24
22	24	21	24
22	23	21	21
23	23	23	22
21	20	23	24
20	23	23	24
20	24	24	20
25	25	23	22
23	20	22	23
24	22	25	24
23	24	24	21
25	21	25	20
21	25	20	22
21	25	23	23
25	24	22	24
25	22	22	22
25	24	23	25
22	23	24	24
22	25	23	21
22	24	25	23
24	22	20	24
21	20	23	21
21	20	20	20
21	21	25	22
21	20	23	22
20	24	21	24
23	21	24	23
24	23	24	25
25	20	25	25
22	24	22	21
20	22	25	24
24	22	24	21
22	20	21	22
21	25	25	24
23	21	20	22
23	20	21	24
23	25	23	25
21	21	25	20
20	22	21	21
25	24	24	23

续表

Swimming Time
25
24
25
24
23
24
20
23
21
23
20
20
24
24
25
24
23
25
21
25
21
21
20
21
23
23
25
21
22
20
21
20
25
21
23
25
20
21
24
24
20
23

Swimming Time
21
22
22
22
23
23
20
24
22
22
22
20
22
23
22
22
23
25
21
21
24
23
20
22
21
21
23
23
20
21
21
20
24
24
21
23
25
20
20
22
22
21

Swimming Time
23
22
24
20
21
22
22
21
22
23
20
20
23
22
22
25
20
24
23
20
24
20
25
24
20
21
20
22
20
22
25
22
24
22
25
22
22
20
23
24
21
22

Swimming Time
24
24
25
20
22
22
24
24
24
25
23
25
23
21
20
21
24
25
25
22
24
24
25
23
25
23
24
25
22
21
23
22
25
24
23
21
23
22
23
23
25
21
22

第 10 章数据集 3

Old Chip Ratings
9.6
6.0
2.0
6.0
5.0
6.9

Old Chip Ratings
7.9
6.6
9.0
8.9
1.0
5.9

Old Chip Ratings
7.3
0.2
3.1
6.5
2.6
4.7

Old Chip Ratings
2.8
2.5
2.5
4.5
0.1
5.8

续表

Old Chip Ratings	Old Chip Ratings	Old Chip Ratings	Old Chip Ratings
4.7	1.8	7.1	2.6
9.7	6.2	3.7	5.1
3.1	7.3	0.9	5.3
5.6	7.7	0.3	5.6
3.8	9.2	5.4	6.6
6.4	5.3	9.1	8.7
7.4	6.7	7.7	3.2
5.2	5.0	6.7	6.7
5.2	8.1	7.7	4.5
0.7	7.9	2.8	6.5
8.1	8.4	1.1	5.3
0.9	7.9	5.3	5.9
0.1	2.3	2.3	4.3
0.3	6.8	5.6	7.9
7.6	1.4	0.9	7.3
2.0	4.8	5.9	6.0
7.9	6.5	2.7	9.6
7.0	7.5	8.9	1.2
2.3	7.8	2.7	1.7
9.8	8.3	4.1	8.1
3.3	5.3	3.7	4.9
9.9	3.2	9.6	5.7
4.8	3.3	0.5	7.8
6.9	9.5	2.4	4.2
0.3	4.4	0.9	1.4
7.6	2.2	8.8	10.0
3.8	3.3	4.5	2.0
6.9	8.7	8.7	9.5
3.1	1.1	8.8	9.0
7.4	6.6	8.0	2.4
3.9	7.1	2.3	9.7
9.8	2.3	2.4	5.8
6.2	4.0	2.0	6.9
2.7	6.3	5.1	8.2
0.2	9.8	9.8	9.5
7.3	9.6	0.8	9.1
9.4	7.3	2.9	4.0
3.1	0.7	2.9	1.3
0.8	3.3	9.5	6.1
3.7	9.4	8.2	6.7
5.6	6.2	3.4	8.7
2.9	3.5	9.6	7.3
4.3	3.4	8.3	7.5
4.3	2.0	4.6	9.9
2.6	8.4	1.7	7.9
1.4	5.3	1.6	8.9
9.9	5.0	7.5	7.1
1.0	1.0	7.1	4.9
5.9	0.6	7.6	3.4
2.7	9.4	2.2	0.3
9.3	0.9	1.3	9.5
3.4	1.9	6.0	2.9

续表

Old Chip Ratings	Old Chip Ratings	Old Chip Ratings	Old Chip Ratings
8.7	4.9	9.2	3.7
6.5	9.7	2.4	1.4
3.8	8.9	7.1	6.8
7.6	10.0	8.1	5.9
5.4	0.0	5.1	8.5
7.3	5.6	5.0	8.5
1.0	3.0	2.2	4.5
0.6	4.5	4.8	2.6
5.7	6.0	4.6	1.4
9.7	5.3	8.4	9.9
9.8	4.2	9.8	8.1
2.2	2.1	6.3	5.3
2.8	6.9	8.4	9.3
9.8	6.5	6.3	5.4
2.5	3.0	7.1	3.3
5.0	1.7	8.5	0.5
3.4	0.8	2.6	1.3
1.8	8.5	0.5	1.5
0.5	0.5	4.1	5.6
8.6	5.5	8.2	8.7
0.2	2.2	8.0	6.1
8.9	1.0	0.2	8.3
1.3	1.0	1.5	8.1
2.8	9.1	5.8	5.6
6.6	9.0	9.0	10.0
4.6	6.9	2.4	8.6
5.7	4.4	5.4	8.8
6.3	8.5	2.6	4.5
4.9	2.3	6.1	5.8
8.7	2.5	0.9	7.7
3.6	5.2	0.8	3.3
1.1	5.8	1.2	3.5
1.8	5.1	9.1	7.4
9.6	4.6	0.8	9.6
9.0	8.9	8.3	1.8
3.8	4.0	7.0	7.8
3.1	6.6	0.8	9.7
6.7	1.4	1.6	1.6
6.8	7.0	7.5	9.5
2.9	5.9	1.1	5.4
6.8	6.6	9.2	3.5
6.9	6.9	9.6	0.5
6.1	9.3	3.9	9.6
2.0	4.1	1.9	3.6
2.3	5.9	3.3	9.0
3.4	2.6	5.0	7.4
5.6	2.0	1.7	2.8
4.0	6.7	1.0	6.0
5.3	0.9	2.1	8.3
1.8	4.3	4.8	4.5

续表

Old Chip Ratings	Old Chip Ratings	Old Chip Ratings	Old Chip Ratings
4.0	5.6	6.7	3.7
0.6	6.5	9.0	9.0
9.2	0.5	6.0	8.9
3.9	1.8	8.7	0.6
0.3	7.6	6.3	3.1
2.1	1.2	3.6	5.9
7.8	3.5	0.2	9.4
3.2	2.0	9.8	2.0
9.2	6.1	4.1	7.3
0.3	8.6	1.4	7.4
6.8	2.1	5.4	8.2
0.1	4.1	6.5	5.7
8.4	9.2	6.2	0.1
7.2	3.0	4.8	8.6
7.7	3.7	2.1	7.0
3.8	6.5	8.2	3.6
1.4	0.3	5.3	1.1

第 11 章数据集 1

Males	Females
9	3
8	5
4	1
9	2
3	6
8	4
10	3
8	6
9	7
8	9
10	7
7	3
6	7
12	6
	8
	8

第 11 章数据集 2

Urban	Rural
6.5	7.9
9.9	4.3
6.8	6.8
4.8	6.5
4	3.3
5.3	13.2
8	9.3
4.2	1.3
7	6.7
6	5.3
9.3	2.4
6.4	4.3
9	1
5.6	3.5
6.6	
5	

第 11 章数据集 3

Sales	Main Street Store	Mall Store
Week 1	$ 3 453	$ 2 542
Week 2	$ 5 435	$ 3 221
Week 3	$ 3 656	$ 1 423
Week 4	$ 4 543	$ 1 656
Week 5	$ 4 543	$ 4 324
Week 6	$ 1 232	$ 3 234
Week 7	$ 4 543	$ 2 312
Week 8	$ 5 643	$ 1 324
Week 9	$ 4 354	$ 2 178
Week 10	$ 6 342	$ 5 468
Week 11	$ 4 355	$ 2 432
Week 12	$ 3 232	$ 2 123
Week 13	$ 6 532	$ 1 543
Week 14	$ 3 234	$ 1 121
Week 15	$ 3 545	$ 4 231

第 12 章数据集 1

Before Recycling	After Recycling	Before Recycling	After Recycling
20	23	23	22
6	8	33	35
12	11	44	41
34	35	65	56
55	57	43	34
43	76	54	51
54	54	22	21
24	26	34	31
33	35	32	33
21	26	44	38
34	28	17	15
33	31	28	17
54	56		

第 12 章数据集 2

Before Intervention	After Intervention	Before Intervention	After Intervention
1.3	6.5	9.0	8.4
2.5	8.7	7.6	6.4
2.3	9.8	4.5	7.2
8.1	10.2	1.1	5.8
5.0	7.9	5.6	6.9
7.0	6.5	6.2	5.9
7.5	8.7	7.0	7.6
5.2	7.9	6.9	7.8
4.4	8.7	5.6	7.3
7.6	9.1	5.2	4.6

第 12 章数据集 3

Before Treatment	After Treatment	Before Treatment	After Treatment
45	46	41	38
46	44	47	31
32	47	41	22
34	42	32	36
33	45	22	36
21	32	34	27
23	36	36	41
41	43	19	44
27	24	23	32
38	41	22	32

第 13 章数据集 1

<15 Hours Practice	15~25 Hours Practice	More Than 25 Hours Practice
58.7	64.4	68
55.3	55.8	65.9
61.8	58.7	54.7
49.5	54.7	53.6
64.5	52.7	58.7
61	67.8	58.7
65.7	61.6	65.7

续表

<15 Hours Practice	15~25 Hours Practice	More Than 25 Hours Practice
51.4	58.7	66.5
53.6	54.6	56.7
59	51.5	55.4
	54.7	51.5
	61.4	54.8
	56.9	57.2

第 13 章数据集 2

	High School 1	High School 2	High School 3	High School 4	High School 5
2003	67	82	94	65	88
2004	68	87	78	65	87
2005	65	83	81	45	86
2006	68	73	76	57	88
2007	67	77	75	68	89
2008	71	74	81	76	87
2009	78	76	79	77	81
2010	76	78	89	72	78
2011	72	76	76	69	89
2012	77	86	77	58	87

第 14 章数据集 1

	High Impact	Low Impact		High Impact	Low Impact
Male	76	88	Female	65	65
	78	76		90	67
	76	76		65	67
	76	76		90	87
	76	56		65	78
	74	76		90	56
	74	76		90	54
	76	98		79	56
	76	88		70	54
	55	78		90	56

第 14 章数据集 2

	Treatment 1	Treatment 2	Treatment 3
Severity 1	6	6	2
	6	5	1
	7	4	3
	7	5	4
	7	4	5
	6	3	4
	5	3	3
	6	3	3
	7	4	3
	8	5	4
	7	5	5
	6	5	3
	5	6	1
	6	6	2
	7	7	4

续表

	Treatment 1	Treatment 2	Treatment 3
	8	6	3
	9	5	5
	8	7	4
	7	6	2
	7	8	3
Severity 2	7	7	4
	8	5	5
	8	4	6
	9	3	5
	8	4	4
	7	5	4
	6	4	6
	6	4	5
	6	3	4
	7	3	2
	7	4	1
	6	5	3
	7	6	2
	8	7	2
	8	7	3
	8	6	4
	9	5	3
	0	4	2
	9	4	2
	8	5	1

第 14 章数据集 3

	Before Intervention	After Intervention
Nutrition Level 1	53	3
	7	7
	9	5
	7	4
	6	6
	7	5
	8	7
	8	6
	9	6
	8	5
Nutrition Level 2	10	4
	10	6
	6	7
	8	8
	7	7
	9	6
	5	9
	6	5
	8	7
	9	8
Nutrition Level 3	8	7
	7	6
	6	5
	6	8
	7	9
	9	7
	8	6
	4	5
	7	6
	8	7

第 15 章数据集 1

Motivation	GPA	Motivation	GPA
1	3.4	6	2.6
6	3.4	7	2.5
2	2.5	7	2.8
7	3.1	2	1.8
5	2.8	9	3.7
4	2.6	8	3.1
3	2.1	8	2.5
1	1.6	7	2.4
8	3.1	6	2.1
6	2.6	9	4
5	3.2	7	3.9
6	3.1	8	3.1
5	3.2	7	3.3
5	2.7	8	3
6	2.8	9	2

第 15 章数据集 2

Number of Books	Educational Level
68	11
345	15
276	16
756	12
43	6
546	14
58	6
187	14
286	9
93	8
376	11
623	18
876	12
28	4
289	15

第 16 章数据集 1

Time	*Correct*
14.5	5
13.4	7
12.7	6
16.4	2
21.0	4
13.9	3
17.3	12
12.5	5
16.7	4
22.7	3

第 16 章数据集 2

Health	Sweets	Gender	BMI
92	1	1	21
76	4	2	23
84	4	2	25
56	6	1	31
73	6	2	22
48	8	2	28
98	2	1	22
96	4	1	23
70	6	1	26
27	8	2	35
63	7	1	27
49	5	2	29
72	5	2	26
57	8	2	27
66	7	1	25
98	2	2	19
69	7	1	32
84	3	1	27
78	8	1	28
55	6	1	26

附录 D 练习题参考答案

1A 小节

1a-1d

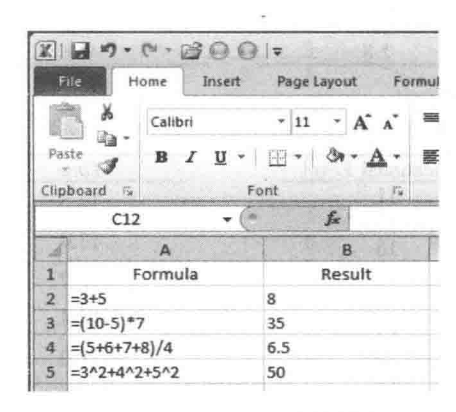


图 1A.16 问题 1a-1d 的答案

第 2 章

1.手动计算

	Score1	Score2	Score3
Mean	5.6	27.6	144.3
Median	5.5	25.0	149.5
Mode	5	25.34	154

图 2.17 所示是 Excel 的电子工作表。要记住数据集 Score 2 有两个众数(是 34 和 25),但是 Excel 只显示一个,在这里显示了 34。

2.你可以看到图 2.17 所示结果。你可以使用众数、中位数,但是均值(mean)(在 Excel 中使用 AVERAGE 一词)是最好的测量值,因为变量(举手的次数)不是分类变量也没有极值出现,所以均值是最精确的。

3.你写的报告可能是这样的:

和往常一样,Chicken Littles(众数)销售量最高。食品销售总额是 303 美元,特惠食品的平均销售价格是 2.55 美元。

4.如果有极值就使用中位数,不然极值就会给均值带来偏差。例如分析收入状况时,收入中位数比均值更适合。因为收入变化很大,你需要集中趋势量数对极值不敏感。另一种情况是你有一个极值或异常值,如你测量一组青少年 100 码跑速度,其中总会有一两个人的速度非常快(或者慢)。

	A	B	C	D
1		Score 1	Score 2	Score 3
2		3	34	154
3		7	54	167
4		5	17	132
5		4	26	145
6		5	34	154
7		6	25	145
8		7	14	113
9		8	24	156
10		6	25	154
11		5	23	123
12	Mean	5.6	27.6	144.3
13	Median	5.5	25.0	149.5
14	Mode	5.0	34.0	154.0

图 2.17 使用 Excel 计算题 1 的数据的均值、中位数和众数

5.下面是 5 个例子。

高中停车场最流行的车,以车名测量。

高中会议讨论的最不喜欢的足球选手,以喜欢/不喜欢测量。

最不成功的测试结果,以五次测试不通过的学生人数测量(通过/没通过)。

最不喜欢的沙拉牌子,以喜欢不喜欢测量。

最喜欢的披萨店,以喜欢不喜欢测量。

6.图 2.18 所示是每月销售均值。我们在一个单元格插入 AVERAGE 函数,然后在第二行复制,然后拷贝到 E 列。

	A	B	C	D	E
1	Toy	July Sales	August Sales	September Sales	Average Sales
2	Slammer	$12,345.00	$ 14,453.00	$ 15,435.00	$ 14,077.67
3	Radar Zinger	$31,454.00	$ 34,567.00	$ 29,678.00	$ 31,899.67
4	Potato Gun	$ 3,253.00	$ 3,121.00	$ 5,131.00	$ 3,835.00
5	Average Sales	$15,684.00	$ 17,380.33	$ 16,748.00	$ 16,604.11

图 2.18 Havefun.com 第三季度每月玩具销售均值

7.如图 2.19 所示,每种玩具的销售中位数在单元格 E2 到 E4。猜猜使用了什么函数?

	A	B	C	D	E
1	Toy	July Sales	August Sales	September Sales	Median Sales
2	Slammer	$12, 345	$14, 453	$15, 435	$14, 453
3	Radar Zinger	$31, 454	$34, 567	$29, 678	$31, 454
4	Potato Gun	$3, 253	$3, 121	$5, 131	$3, 253
5	Median Sales	$12, 345	$14, 453	$15, 435	$14, 453

图 2.19 Havefun.com 第三季度每月玩具销售中位数

8.你会选择使用中位数,因为中位数对极值不敏感。

9.

	12/1~12/7	12/8~12/15	12/16~12/23
Mean	19.25	15.50	17.25
Median	13.50	13.00	17.00

至少对 12/1 到 12/7 周来说,中位数是更有代表性的集中趋势量数,因为 15~19 岁年龄组存在极值 38。对所有的周来说,均值看起来更合适。

第3章

1.极差是最方便的离散趋势量数，只需要用一个数（最大值）减去另一个数（最小值）。极差没有考虑一个数据分布中最大值和最小值之间的其余数值，不太精确。如果你只是大概地（不是很精确）估计数据分布的变异性，就可以选择使用极差。

2.

最高成绩	最低成绩	包含极差	不包含极差
7	6	2	1
89	45	45	44
34	17	18	17
15	2	14	13
1	1	1	0

3.极差等于30。样本标准差无偏估计值等于10.19，有偏估计值等于9.60。两者之间的差异是由于一个使用样本规模8作除数（无偏估计），另一个使用样本规模9作除数（有偏估计）。方差的无偏估计是103.78，有偏估计是92.25。

4.标准差的无偏估计是12.10，方差的无偏估计是146.23。

5.标准差的有偏估计是11.47，方差的有偏估计是131.61。

6.无偏估计始终比较大有两个原因。首先，计算无偏估计的公式的分母是 $n-1$ 而不是 n，导致分数的值始终比较大。更重要的是因为我们想保持保守，无偏估计（基于样本计算，不能给我们足够的和总体一样多的信息）必须比较大，这样我们就能弥补这样的事实，即我们只是分析总体中的一小部分。

7.图3.6所示是数据和计算所得的标准差和方差。我们使用无偏函数（STDEV.S and VAR.S）和有偏函数（STDEV.P and VAR.P）计算。事实上，你当然也可以只看公式栏，查看单元格内容来看函数的命令句形式。

8.首先，事实上（我们之前使用分析工具库中的描述统计选项，这是最有效的方式）早航班的平均乘客人数是244（实际是244.33），晚航班的平均乘客人数是296（实际是296.5）。

早航班的标准差是61.74，晚航班的标准差是47.35。

你可以计算任何一个集中趋势量数和变异性量数，并且进行解释。例如，乘坐晚航班的人数比早航班的人数多（这可能意味着销售了更多的单程机票等），而且如果是晚航班，人数似乎更具有稳定性（标准差比较小）。（还可以计算更多的描述统计值，也非常有意思，比如进行城市不同时间航班人数的比较。）

9.图3.7是经过整理的使用分析工具库描述统计选项所得结果，表明没有干预的一组的平均值（17.80），略高于干预组的平均值（17.60）。标准差也接近，没有干预组是4.68，干预组是5.60。

	A	B	C
1		Height	Weight
2		53	156
3		46	131
4		54	123
5		44	142
6		56	156
7		76	171
8		87	143
9		65	135
10		45	138
11		44	114
12		57	154
13		68	166
14		65	153
15		66	140
16		54	143
17		66	156
18		51	173
19		58	143
20		49	161
21		48	131
22	STDEV.S	11.44	15.65
23	VAR.S	130.78	245.00
24	STDEV.P	11.15	15.26
25	VAR.P	124.24	232.75

图 3.6 计算无偏和有偏标准差和方差

M26

	A	B	C	D	E	F	G
1	No Intervention	Intervention					
2	18	17		*No Intervention*		*Intervention*	
3	16	15					
4	17	21		Mean	17.80	Mean	17.60
5	15	24		Standard Error	1.21	Standard Error	1.45
6	16	12		Median	18	Median	17
7	14	16		Mode	16	Mode	21
8	22	14		Standard Deviation	4.68	Standard Deviation	5.60
9	21	25		Sample Variance	21.89	Sample Variance	31.40
10	7	19		Kurtosis	0.45	Kurtosis	-1.29
11	24	24		Skewness	-0.73	Skewness	-0.09
12	23	21		Range	17	Range	16
13	19	25		Minimum	7	Minimum	9
14	12	9		Maximum	24	Maximum	25
15	20	9		Sum	267	Sum	264
16	23	13		Count	15	Count	15

图 3.7 阅读干预和没有干预比较:分析工具库结果

第 4 章

1a.下面就是频数分布表。

组 距	频 数	组 距	频 数
55—59	7	25-29	5
50—54	5	20-24	4
45—49	5	15-19	5
40—44	7	10-14	3
35—39	2	5-9	3
30—34	3	0-4	1

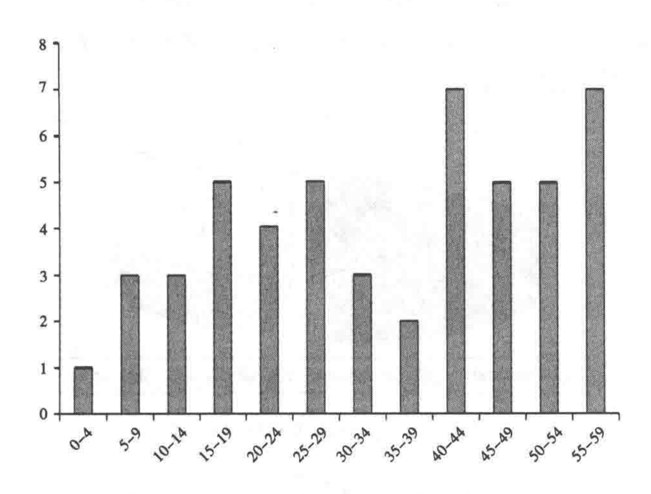


图 4.31 第 4 章数据集 1 数据的直方图

1b.我们设定的组距是 5,这个标准符合我们本章讨论的设定组距的标准。我们想要的组距范围是 2、5、10 或者 20,我们想要使用 10~20 个组距覆盖数据。这个例子中我们选择组距 5,使用 12 个组距。

1c.这是负偏度分布，因为均值小于中位数。

2a.线图，线图可以表示随时间发生的变化。

2b.饼图，可以直观看出整体不同部分所占的相对比例。

2c.条形图，对于展示分类数据或者定类数据非常有优势。

3a.饼图，显示比例。

3b.线图，展示随时间发生的变化。

3c.条形图或者柱状图，展示分类数据。

3d.线图，展示重复刺激发生的变化。

3e. 条形图或者柱状图，展示分类数据。

4.自己来完成。

5.你可以在图 4.32 看到结果。这是非常有趣的图形展示，也说明了在某些时候可以使用图像有效地传递意义。不过，这样的图像表示不像条形图、柱状图或折线图那样严谨。

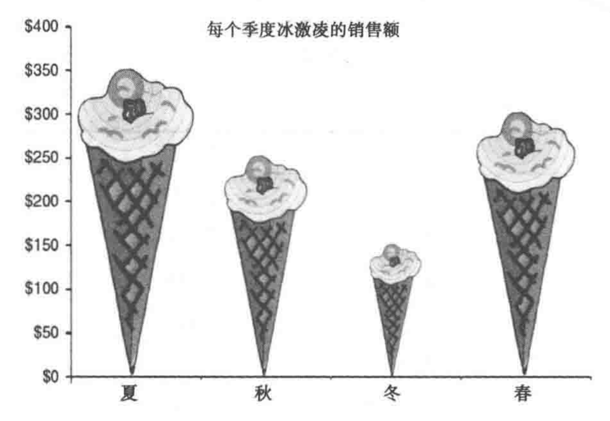


图 4.32 图例——使用图像代替条形

6.我们使用 Excel 和图表编辑器完成这个饼图，结果是既难看又没有太多信息。图 4.33是弗兰克斯坦因博士修改后的饼图。

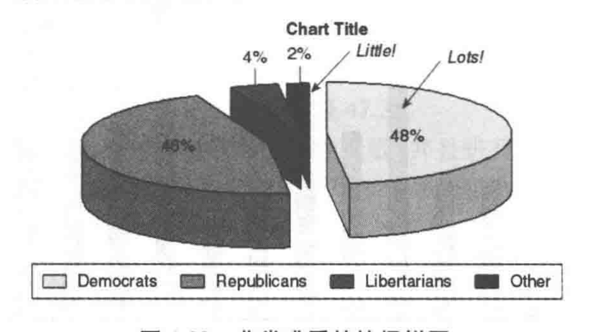


图 4.33 非常难看的垃圾饼图

第 5 章

1a. $r = 0.596$。

1b.依据 1a 的答案，你已经知道相关是正向的。但是依据如图 5.13 所示的散点图，你可以预测这样的结论（即使实际并不知道相关系数的正负号），因为数据点集合本身是从

图的左下角到右上角,可以判定是正斜率。

Correct	Attitude
17	94
13	73
12	59
15	80
16	93
14	85
16	66
16	79
18	77
19	91

Attitude by Number Correct

图 5.13　正确回答的问题数量和对待考试的态度的散点图表明正向的关系

2a. $r=0.269$。

2b.依据表 5.2,这个规模的相关强度是弱相关。决定系数是 0.269 2,即方差的 7.2%(0. 072)可以得到解释。主观分析(弱相关)和客观数值(可解释方差的 7.2%)一致。

3.相关系数是 0.45,这意味着对医生的培训越多,结果就越好(记住 1 是好的,10 不是),也就是说医生获得的培训越少,结果越坏。将结果 0.45 平方得 0.20,或者说一个变量 20%的变异性可以被另一个变量解释。这只是个例子,在真实的生活中,也许并非如此。

4.如图 5.14 所示的是相关矩阵,就如你看到的,洗手效果和洗手次数之间的相关不是很强(从技术上说,不存在相关),洗手效果和感染数量之间甚至是弱相关。但是,在洗手次数和感染之间强相关,这就意味着洗手次数比洗手质量更能减少感染。洗手效果和感染数量之间的关系并不比公鸡打鸣和太阳升起之间的关系更直接。它们可能是同时发生,但是并不是一个导致另一个发生。

Wash	Number	Infect
2	4	1
4	2	2
2	4	2
3	4	3
4	5	1
5	5	1
4	2	5
3	5	4
2	4	2

	Wash	*Number*	*Infect*
Wash	1		
Number	-0.07625	1	
Infect	0.026958	-0.42931	1

图 5.14　相关矩阵:洗手效果、吸收频次和医院感染数量

5a.0.8。

5b.非常强的相关。

5c.1-0.64,即 0.36(36%)。

6.你应该使用卡方系数检验种族和政党背景的关系,因为这两个变量是定类的。你应该使用点二列相关系数检验俱乐部成员身份和高中平均成绩(GPA)的关系,因为这两个变量的属性一个是定类的(俱乐部成员身份),另一个是定距的(GPA)。

7.如果变量之间有共享的部分,那么变量之间就相关。例如受教育水平和终生收入、体重和血糖水平、外科手术满意程度和期望。但是你不能得出结论认为一个变量的变化会引起另一个变量的变化,因为相关反映的是关联而不是因果关系。许多事情“非常巧合地”发生——就如冰激凌和犯罪。是否还记得?

8.规则制定人员一定要非常谨慎，但是他/她也还是会出错。这项声明暗示青少年怀孕是辍学导致的（我们知道事实不是如此）。实际上，这两个变量只是相互关联，这种关系可能正好是他们共享的部分引起的（如家庭收入水平或者其他与早孕、辍学相关的重要的变量）。

9.相关系数是0.64。这意味着预算增加和班级成绩增加之间是正相关（但是要注意的是我们还得检验相关的显著性）。从描述统计的角度来看，这两个变量大概共享36%的方差。

10.

+0.36　　−0.45　　+0.47　　−0.62　　+0.71

第6章

1.自己来完成。

2.如果你对不同时期成绩测试的一致性感兴趣，你可以使用再测信度，例如你在8月也就是学期开始测试孩子的技能，在5月学期末再次测试。至于平行信度，你想给两个青少年群体进行相同现象的测试，测试形式一种是通过纸和笔书写，另一种是使用文字处理软件。比较两种形式并确保评估工具是准确和真实的。

3.因为需要使用再测信度，我们使用CORREL函数计算相关系数是0.84，这个值足够大，表明具备很高的信度。

4.再测信度低于0.14，意味着信度很低。

5.一项测试是可信的，但是不能测量要测量的内容，那么这项测试就是有信度的，但是没有效度。测试具备效度，同时要具备持续性（也就是要有信度）。如果测试没有信度，怎么可以说测试了要测试的内容？我们不可以这样说。

6.这是涉及上百万的问题。如果没有使用既有信度又有效度的测量工具，如何知道假设能够得到支持（或不能得到支持）？也许是测量工具的质量本身得到正面或者负面的结果。

7.同步效度评价一项测试是否很好地反映了特定的结构、性质、特点或能力表现，而预测效度预测将来的表现。例如同步效度可以判断一项测试如何准确地评价恰当的举重技能，而预测效度是评价一个问卷调查是否能很好地预测举重竞赛的成功。

第7章

1.问题1和问题2依据你自己的兴趣决定。虽然没有正确答案，但是很多是错误的。要很好地回答这两个问题，你可以尽力查找与你的专业兴趣相关的领域，以及你愿意去读的文章。

3a.零假设：依据注意时间观察量表的测量，注意时间短的学生和注意时间长的学生不专心听课行为的频率相同。

有方向研究假设：依据注意时间观察量表的测量，注意时间短的学生比注意时间长的学生不专心听课行为的频率要高。

无方向假设：依据注意时间观察量表的测量，注意时间短的学生不专心听课行为的频率不同于注意时间长的学生。

3b.零假设：婚姻质量和夫妇双方与他们兄弟姐妹关系的好坏之间没有关系。

有方向研究假设：婚姻质量和夫妇双方与他们兄弟姐妹关系的好坏之间有正向关系。

无方向假设:婚姻质量和夫妇双方与他们兄弟姐妹关系的好坏有关。

3c.零假设:结合传统心理治疗的药物治疗和单独的传统心理治疗对治疗厌食症有相同的效果。

有方向研究假设:结合传统心理治疗的药物治疗比单独的传统心理治疗对治疗厌食症更有效。

无方向假设:结合传统心理治疗的药物治疗和单独的传统心理治疗对治疗厌食症具有不同的效果。

3d.零假设:参与早期阅读项目对后期的阅读成绩没有影响。

有方向研究假设:参与早期阅读干预项目的孩子进入六年级后阅读成绩将要高于没有参与阅读项目的孩子。

无方向假设:参与早期阅读干预项目的孩子进入六年级后阅读成绩将会与没有参与阅读项目的孩子不同。

4.这应该由你自己来完成,但是要记住如果研究假设越完整越明确,就越接近我们在讨论中特别确定的标准。简单和明确——这就是我们想要的。

5.零假设是两个测量之间相等的陈述。零假设最重要的目的之一是成为任何研究的起点,也就是"零"陈述是一个等量关系。也就是说,我们在对变量之间关系完全不了解的情况下,可以得出的最可靠的结论就是它们的值相等。零假设与研究假设的不同之处就是研究假设是不相等关系的陈述。

6.假定你正在研究一个问题(接着会形成一个假设),你对于结果所知甚少(这就是你为什么问这个问题,并进行检验的原因),那么零假设就是很好的开始,因为零假设是一个相等的命题陈述,这是最基本的假定。"假定对于正在研究的关系没有更多的信息,我只能从零开始,也就是我知道的很少。"零假设是完美的、无偏的、客观的起点,因为任何事物都假定是相等的,除非可以证明是不相等的。如果假定其他关系,就是假定你在开始时有偏差,这不是我们开始一项研究时想要的陈述。

7.

零假设	研究假设
a.零假设:大学生运动员食品支出的金额与其他大学生没有差别。 $H_0: \mu_{us} = \mu_{sa}$	a.研究假设:大学生运动员食品支出金额比其他大学生多。 $H_1: X_{us} < X_{sa}$
b.零假设:小白鼠和小棕鼠走出迷宫花费的平均时间没有差别。 $H_0: \mu_W = \mu_B$	b.研究假设:小白鼠和小棕鼠走出迷宫花费的平均时间有差别。 $H_1: X_W \neq X_B$
c.零假设:药物 A 的效果和药物 B 的效果没有不同。 $H_0: \mu_A = \mu_B$	c.研究假设:药物 A 的效果比药物 B 的效果强。 $H_1: X_A > X_B$
d.零假设:用方法 1 完成任务花费的时间与方法 2 花费的时间相同。 $H_0: \mu_1 = \mu_2$	d.研究假设:用方法 1 完成任务花费的时间与方法 2 花费的时间不相同。 $H_1: X_1 \neq X_2$

第8章

1a.对于正态曲线，均值、中位数和众数相等；曲线是以均值为中心对称的；曲线尾是渐近的。

1b.例如身高、体重、智力或问题解决能力。

2.因为 z 值是基于不同数据分布的离散度计算的，所以是标准化的值（可以与其他同类型数值比较）。因为 z 值是测量均值和横轴上其他数据点之间的距离（不论数据分布在均值和标准差上的具体差异如何），使用相同的单位（标准差单位），因而不同数据分布能够相互比较。

3.z 值或其他任何重要的标准分都允许你使用相同的单位进行不同分布的数值的比较。换句话说，最重要的特点是，一个分布的特征不重要，而来自不同分布的标准分是可比较的。一个好的例子如下，如果你想比较来自不同测试情况下的原始数值，要把这些数值转换为标准分来进行比较。

4a.$z=(55-50)/5=+1.00$

4b.$z=(50-50)/5=0$

4c.$z=(60-50)/5=+2.00$

4d.$z=(57.5-50)/5=+1.7$

4e.$z=(46-50)/5=-0.8$

5.如果均值是 40，标准差是 5，与原始数值相对应的 z 值是 1.5，那么原始数值是

$X = 5(1.5) + 40 = 47.5$

相应的 t 值是 $t=50+10(z)=65$

6a.一个数值落在原始数值 70 和 80 之间的概率是.564 6。原始数值 70 的 z 值是 −0.78，80 的 z 值是0.78。均值与 z 值0.78之间曲线覆盖的面积是 28.23%。这两个 z 值之间曲线覆盖的面积是 28.32%×2，或 56.46%。

6b.一个数值落在原始数值 80 以上的概率是 0.217 7。原始数值 80 的 z 值是 0.78。均值与 z 值 0.78 之间曲线覆盖的面积是 28.23%。z 值 0.78 之下曲线覆盖的面积是0.50+0.282 3，或 0.782 3。曲线覆盖的总面积 1 和 0.782 3 之间的差是 0.217 7，或 21.77%。

6c.一个数值落在原始数值 81 和 83 之间的概率是 0.068。原始数值 81 的 z 值是0.94，83 的 z 值是 1.25。均值与 z 值 0.94 之间曲线覆盖的面积是 32.64%。均值与 z 值1.25之间曲线覆盖的面积是 39.44%。这两个 z 值之间曲线覆盖的面积是 0.394 4−0.326 4=0.068，或者 6.8%。

6d.一个数值落在原始数值 63 以下的概率是 0.03。原始数值 63 的 z 值是−1.88。均值与 z 值−1.88 之间曲线覆盖的面积是 46.99%。z 值−1.88 之下曲线覆盖的面积是 1−(0.50+0.469 9)= 0.03，或者 3%。

7.这比你想的要容易。只要输入下面的公式，可以在任何单元格使用 NORM.S.DIST 函数，然后按回车键。

=NORM.S.DIST(2,TRUE)−NORM.S.DIST(1,TRUE)

神奇的答案就是 0.135 905 122。

8.如图 8.11 所示，可以看到结果，所示的公式从 B2 复制到了 B11。

9.我们知道代表分布 90%以上的 z 值是 1.29，我们把这个值代入公式 $X = zs + \overline{X}$，也就是 $X=1.29(5.5)+78=85.095$。我们进行四舍五入，也就是说杰克需要至少 85 分才能拿到证书。

10.理解中心极限定理在推论统计中所扮演的角色，是理解总体不具备正态分布所有

特点,但是还可以被看作正态分布。样本的多重选择将会导致这些样本的数值类似一个正态分布。

11.这不是因为原始数值属于不同的分布而不能比较。数学测试的班级平均值是40,原始得分是80,这个值不能和写作技巧测试成绩80分进行比较。不同数据分布(就像不同人一样)并不总是可以互相比较。并不是所有的事物(人)可以和其他事物(人)进行比较。

B2 =NORM.S.DIST(A2,TRUE)

	A	B	C	D	E	F
1	z Scores	Probability				
2	1.5	0.933193				
3	2.1	0.982136				
4	0.7	0.758036				
5	0.4	0.655422				
6	2.0	0.977250				
7	1.6	0.945201				
8	0.4	0.655422				
9	0.8	0.788145				
10	0.5	0.691462				
11	1.7	0.955435				

图 8.11 使用 NORMAL.S.DIST 函数计算 z 值得概率

12.下面的表格中填入了 z 值。

	数学		
班级均值	81		
班级标准差	2		
	阅读		
班级均值	87		
班级标准差	10		
		原始值	
	数学成绩	阅读成绩	平均成绩
Noah	85	88	86.5
Talya	87	81	84
		z 值	
	数学成绩	阅读成绩	平均成绩
Noah	2	0.1	1.05
Talya	3	−0.6	1.2

Noah 原始得分的平均值比较高(Noah 是 86.5,Talya 是 84),但是 Talya 的 z 值平均值比较高(Noah 是 1.05,Talya 是 1.2)。要记住我们的问题是相对其他人来说谁是学习较好的学生,这就需要使用标准分(我们使用 z 值)。但是为什么 Talya 是比 Noah 更好的学生?这是因为在变异性最小(数学测试的标准差是 2)的测试中,Talya 的 z 值是 3,这说明她的成绩排在了前列。

第9章

1.显著性概念是研究和使用推论统计的关键,因为显著性(一般使用显著性水平)设定我们可以确信观察到的结果是"真实的"的水平,以及在什么程度上可以将这些结果推论到样本代表的更大的总体。

2a.显著性水平只用于单一的、独立的零假设检验,而不是多元假设检验。

2b.设定错误的概率水平为 0 是不可能的,因为我们可能在零假设为真的情况下拒绝零假设。这样的小概率事件始终存在。

2c.你愿意承担的零假设为真的情况下拒绝零假设的风险水平，与研究结果的意义之间没有关系。你可以得到显著性水平很高但是没有意义的结论，或者第一类错误的概率水平很高（0.10），但却非常有意义的发现。

3.随机因素反映可能拒绝真实零假设的风险水平（第一类错误）。

4a.因为结果的概率水平低于0.05（或者是你随机预期的结果），拒绝零假设，或者说拒绝"关系是显著的"。

4b.因为没有关系所以没有拒绝零假设。换句话说，咖啡的消耗量和平均成绩之间不相关。

4c.因为变量之间没有关系，所以没有拒绝零假设。换句话说，工作时间和工作满意度之间不相关。

5.在0.01的显著水平下，因为检验更稳健，误差或者错误的概率非常小。换句话说，如果结果出现的概率比较小（比如是0.01）而不是比较大（比如是0.05），就更"难"说结果的出现是随机因素导致（零假设）的了。

6.这有点像玩文字游戏，但是就是这样。因为你永远不能直接地检验零假设（记住零假设对应总体，而研究假设对应样本）。根据对样本的检验结果，要么是研究假设成立，"接受研究假设"；要是没有证明表明研究假设成立，只好是暂时"无法拒绝零假设"。

7.临界值是一个分界点。在这个分界点，比较小的极值因为随机因素而发生（如果零假设是"正确的"），较大的极值表明结果是由于随机因素之外的因素引起。

8.实际值是应用统计检验来评估某个结果随机发生的概率值。

9.下面就是流程图的步骤，从最上面开始。

两个群体是独立的；一个是样本，另一个是总体。

检验两个群体在一个或者多个变量上的差异。

不同的参与者接受测试。

两个群体接受测试。

需要独立样本的 t 检验。

10a.曲线下的面积表示结果不可能是我们预期的随机因素导致的，而更可能是由于处理变量导致的。

10b.因为检验变得更稳健，这个区域变得更小，而且定义这个区域的线条需要向右移动。那么错误（第一类错误）发生的可能就更小，因此这个区域更小。

第10章

1.如果比较样本均值和总体的均值，单样本 Z 检验是合适的，而且零假设是两者之间没有差异。

2.非常有趣，Z 检验中的 z 值是标准分，与简单的 z 值一样。与简单的 z 值或者其他标准分一样是由于这样的事实，在不同分布之间可以进行比较，而且分母是均值的标准误。这种形式是与分布的所有均值的均值偏差，而不是仅仅与一个均值的偏差。但是与简单的 z 值类似，这是依据均值标准化的值，Z 检验的值也可以在不同分布之间比较。这就是为什么正态曲线下的面积如此有用——提供了不同事件的可比性。

3a. $H_1:\overline{X}_{Chocolate\ Only\ Die} \neq \mu_{Chocolate\ Only\ Die}$

3b. $H_1:\overline{X}_{Rate\ Per\ Thousand} \neq \mu_{Average\ Rate\ Last\ 50\ Seasons}$

3c. $H_1:\overline{X}_{Monthly\ Costs} \neq \mu_{Monthly\ Cost\ Over\ Past\ 20\ Years}$

4.下面就是计算结果。

$$SEM = \frac{2.35}{\sqrt{15}} = 0.60 \qquad z = \frac{16-15}{0.60} = 1.66$$

计算所得的 z 值是 1.66,查阅附录 B 中的数据表,对我们来说这个值不是足够大,不能得出结论认为巴特菲尔德地区的流感发生数量与整个州的发生数量不同。不要忘记经常洗手哦。

5.与通常一样,布兰达是明星。她的 55 码游泳时长发生在这个 1 000 人的新总体中的概率是 0.006(这是 Z.TEST 函数返回的值),显然,与新片区的人相比,她游得非常快(是在极值区域)。

6.新薯片(打分平均值是 5.1)的概率是 0.328 9。新薯片的口味肯定不糟糕,概率水平很高,这意味着新薯片的味道和真的薯片没有差别。那么就开拓新薯片的市场吧!

第 11 章

1.男孩的举手次数的均值是 7.93,女孩的均值是 5.31。统计值 t 值是 3.006,在0.05的显著水平下单侧检验拒绝零假设的临界值 t 值是 1.701。那么结论是什么?男孩举手的次数显著地高于女孩。

2.现在的问题更有趣。我们有完全相同的数据,但是假设却不同。现在的假设是举手次数是不同的(而不是多或者少),所以需要进行双侧检验。因此使用附录表 B2,在 0.01 的显著水平下双侧检验的临界值是 2.764。实际值 3.006(与分析问题 1 所得结果相同)超过了临界值,就这个假设而言,男生和女生的举手次数不同。因此两项检验比较而言,使用相同的数据得出相同的结论(接受研究假设)情况下单侧检验的结果(见问题 1)不需要和双侧检验结果一样。

3.我们使用 T.TEST 函数,得到的值是 0.254。这就是作为居民态度的两组数据差异的准确概率。猜猜结论是什么?他们的态度完全不同。

4.两个店比较的结论(要记住我们已对结果进行了整理,使得其看起来更简洁)如下。

	Main Street	Mall
均值(Mean)	\$ 4 309.47	\$ 2 608.80
方差(Variance)	1 844 069.981	1 598 119.457
观察数(Observations)	15	15
自由度(df)	28	
t 值(t Stat)	3.55	
单侧检验概率 P(T<=t) one-tail	0.000 6	
双侧检验概率 P(T<=t) two-tail	0.001 3	

因为与两个店的差异 \$ 1 700.67(4 309.47−2 608.80)相对应的概率是 0.001 3,显然两个店是不同的。(记住,这是双侧检验,因为老板只想知道是否有差异而不要求差异的方向。)需要对商店(以及雇员)进行重估。

5.发现仍然有用,但是差异的统计显著性必须始终在所问的问题范围内思考。统计显

著的差异，但是没有有意义的效应量，这表明差异不是由于随机因素导致，可能概念上比能证明自身的统计证据要弱。这也是为什么统计学是一个思考者的运动——数字永远不能做出完全的说明。

第12章

1.独立均值 t 检验用于检验两个不同的参与者群体，每个群体接受一次测试。非独立均值 t 检验用于一个参与者群体，其中每一个参与者接受两次测试。

2.回收项目执行之前的均值是34.44，项目执行之后的均值是34.80。回收量增加了。这25个街区前后回收量的差异是否显著？统计值 t 值是0.234，自由度是24。在0.01的显著水平——研究假设检验确定的显著水平下，差异不显著。结论是回收项目没有促进废纸回收量的增加。

3a.独立均值检验

3b.独立均值检验

3c.非独立均值检验

3d.独立均值检验

3e.非独立均值检验

4.满意水平有所增长，从5.48增长到7.60，相应的 t 值是3.893。相应的差异的概率水平是0.001。也就是社会服务项目在发挥作用。

5.对杰克来说答案是肯定的。t 值是3.34，显著水平是0.03。在培训之前平均销售水平是＄61 849.08，在培训之后是＄83 850.83。聪明的杰克。

6.我们使用分析工具库完成，输出结果如下。

t 检验：配对样本均值		
	干预前	干预后
均值（Mean）	5.48	7.60
方差（Variance）	5.14	2.00
观察数（Observations）	20	20
皮尔森相关系数（Pearson Correlation）	0.19	
假定的均值差异（Hypothesized Mean Difference）	0	
自由度（df）	19.00	
t 值（t Stat）	−3.89	
单侧检验的概率 P（T<=t） one-tail	0.000 49	
单侧检验临界值 t Critical one-tail	1.73	
双侧检验的概率 P（T<=t） two-tail	0.00	
双侧检验的临界值 t Critical two-tail	2.09	

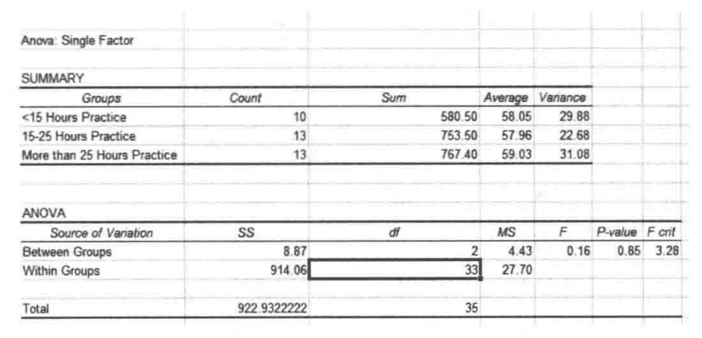
我们的结论是，干预项目在.01的概率水平检验下发挥作用。而且事实上，两个群体之间这个规模的差异由于偶然因素导致的可能性非常小（0.000 49）。

第13章

1.填入案例的表格如下。

设计	分组变量	检验变量
简单方差分析	培训时间分为四个层级——2、4、6 和 8 个小时	打字的准确程度
	三个年龄群体——20 岁、25 岁和 30 岁	力量
	六种工作类型	工作绩效
两因素方差分析	培训的两个层级和性别(2×2 设计)	T 打字的准确程度
	三个年龄群体——5、10、15 岁，兄弟姐妹的数量	社交技能
三因素方差分析	培训的三个层级、性别和收入的三个层级	选举态度
	课程类型(类型 1 和类型 2)，GPA(3.0 以上和以下)，以及活动参与度(参与和不参与)	ACT 成绩

2.图 13.7 所示是分析工具库的部分输出结果，三个的均值分别是 58.05 秒、57.96 秒和 59.03 秒，由于随机因素导致的 F 值($F_{(2,33)}=0.160$)的概率是 0.85，远远大于我们预期的处理因素导致的概率水平。我们的结论是什么？练习的时间对于运动员游泳速度提高没有影响。



Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
<15 Hours Practice	10	580.50	58.05	29.88
15-25 Hours Practice	13	753.50	57.96	22.68
More than 25 Hours Practice	13	767.40	59.03	31.08

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	8.87	2	4.43	0.16	0.85	3.28
Within Groups	914.06	33	27.70			
Total	922.9322222	35				

图 13.7 分析工具库的部分输出结果

3.肯定存在差异。分析结果是 $F_{(2,48)}=63.36$，$p<0.000$，这意味着由广告形式之外因素引起的差异的可能性非常非常低，如果你看三个群体的均值，很明显标签为“Combination(组合色)”的一组最有效。

Group	Average
Color(彩色)	7.06
Black and White(黑与白)	4.53
Combination(组合色)	8.76

4.这五所学校存在明显的差异，因为 F 值是 17.73，相对应的概率水平小于 0.01(实际上是0.000 000 008)。高中学校 5 学生毕业率最高(86.00%)，学校 4 学生毕业率最低(65.20%)。具体见图 13.8。

第 14 章

1.很容易区别，析因方差分析只用于不止一个因素或独立变量的分析。当你假设交互

SUMMARY						
Groups	Count	Sum	Average	Variance		
High School 1	10	709	70.90	21.88		
High School 2	10	792	79.20	24.62		
High School 3	10	806	80.60	38.49		
High School 4	10	652	65.20	94.62		
High School 5	10	860	86.00	13.11		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2733.28	4	683.32	17.73	0.00	2.58
Within Groups	1734.50	45	38.54			
Total	4467.78	49				

图 13.8 一元方差分析

作用存在时，实际上获得结果不是很容易（但是只要你获得结果，你就会真正理解）。

2.这里可以给出许多不同的可能的案例，下面是其中之一。一个处理变量（因素）有三个层级，疾病严重性分为两个层级。

		处理变量		
		药品 1	药品 2	药品 3
疾病严重性	严重			
	轻微			

3.源表类似如下所示。

	ANOVA				
Source of Variation	SS	df	MS	F	p-value
Sample	0.075	1	0.075	0.037	0.848
Column	263.517	2	131.758	64.785	0.000
Interaction	3.150	2	1.575	0.774	0.463
Within	231.850	114	2.034		
Total	498.592	119			

我们的结论是什么？在这个数据集中，疾病严重性没有主效应，处理变量有主效应，两个主要因素之间没有交互效应。

4a.饮食干预的确有效果，非干预群体的平均值是 53.86，干预组的平均值是 71.93，我们也得到干预变量的 F 值是 6.90，在 0.01 风险水平下是显著的。

4b.营养教育项目没有效果：三个层级的平均值没有差异，F 值是 0.82，显著性水平是 0.44。

4c.至于饮食干预和营养教育项目之间的交互效应，没有效果，证据是 F 值是 0.11，第一类错误率是 0.90。

4d.总之，如果你想让人们提高健康饮食的得分，在这种情况下应采取干预项目。不要花钱在什么营养教育项目方面。

第 15 章

1a.自由度是 18（$df=n-2$）、显著水平是 0.01 的情况下拒绝零假设的临界值是 0.516。

速度和力量之间是显著相关,而且相关系数解释方差的 32.15%。

1b.自由度是 78、显著水平是 0.05 的情况下单侧检验拒绝零假设的临界值是 0.183。所以正确的数量和完成时间之间显著相关。因为研究假设变量间关系是负向的,所以使用单侧检验,而且大约解释方差的 20%。

1c.自由度是 48、显著水平是 0.05 的情况下双侧检验拒绝零假设的临界值是 0.273。所以朋友的数量和可能取得的 GPA 之间显著相关,而且相关系数可以解释方差的 13.69%。

2a 与 2b.我们使用分析工具库计算得出动机和平均成绩之间的相关系数为 0.434,显著水平为 0.017 的双侧检验是显著的。图 15.2 给出最后的分析结果。要注意,这里没有给出概率,你必须查阅附录 B4 确定第一类错误水平。

	A	B	C	D	E	F
1	Motivation	GPA				
2	1	3.4			Motivation	GPA
3	6	3.4		Motivation	1	
4	2	2.5		GPA	0.43	1

图 15.2　第 15 章数据集 1 的相关分析工具输出结果

2c.正确。动机水平越高,平均成绩越高。这是正确的,因为两个变量之间存在显著的相关。但是(很重要的"但是")更多的学习并不能引起更高的动机水平,同样的,更高的动机水平并不能导致更多的学习。

3.可以给出的案例很多。这个案例是学习的时间和第一次统计学考试成绩。这两个变量不是因果上的相关。例如,你的同学因为不理解内容,即便学习了几个小时,学习效果还是很差;而有的同学可能在其他课程学习了同样的内容,即使不学习成绩也很好。假设我们强迫一些人在考试前四个晚上坐在书桌前每晚学习 10 个小时,这能保证他或者她得到好的成绩?当然不能。不能因为变量相关就认为其中一个变量的变化就会引起另一个变量的变化。

4.图书的数量和教育水平之间的相关系数是 0.55,自由度是 8,在 0.05 的风险水平下单侧检验是显著的。但是书架上一堆书并不能保证获得博士学位。到现在为止你应该知道为什么:事物之间相关并不(一定)是一个会导致另一个发生。

第 16 章

1a.回归等式是 $Y' = -0.214$(正确数量)$+17.202$

1b.$Y' = -0.214 \times 8 + 17.202 = 15.49$

1c.

时间(Y)	正确数量(X)	Y'	$Y-Y'$
15.5	5	16.13	−1.6
13.4	7	15.70	−2.3
12.7	6	15.92	−3.2
16.4	2	16.77	−0.4
21.0	4	16.35	4.7
13.9	3	16.56	−2.7
17.3	12	15.63	2.7
12.5	5	16.13	−3.6
16.7	4	16.35	0.4
22.7	3	16.56	6.1

2a.其他的估计变量不能和任何一个估计变量相关。只有这些变量相互独立才能够在估计因变量或结果变量时提供独特的贡献。

2b.例如，生活安排（单独过或者与他人一起生活）和获得医疗服务的机会（高、中和低）是估计变量。

2c.阿尔茨海默病的存在$=X_1$（教育水平）$+X_2$（一般健康状况）$+X_3$（生活安排）$+X_4$（获得医疗服务的机会）$+a$。

3.不管你预测什么，估计变量太多会带来很多的成本。首先，要花费时间和金钱去收集如此多的数据；其次，找到在概念定义上与某个变量相关而实际上彼此不相关的如此多的变量是非常困难的。换句话说，你收集的信息可能极大程度上与你已经具有的信息重叠。最大的问题是什么？你如何知道哪一个变量对预测作出贡献？

4.自己来完成。

5a.计算这两个变量之间的相关系数，是0.204。依据第5章的信息，这样的相关系数值比较低。你可以得出的结论是前10年的胜利次数不是一个队伍是否赢得超级杯的很好的估计变量。

5b.许多变量依据属性特征（性别、种族、社会阶级和党派）分类，并且很难依据类似1~100的尺度进行测量。使用分类变量可以给我们很大的弹性空间。在这个案例中，我们研究分类变量，就是赢或输。

5c.其他的估计变量可能是全美球员的数量、教练的输/赢记录和家庭护理。

6.估计的标准误是测量估计模型的效果，也就是估计值（结果变量或者依赖变量）和实际值的差别有多大。因此，如果相关系数小，那么估计一定程度上也不太好，估计标准误就高。

7.很明显这三个变量一起发挥作用，具有的相关系数是0.885，这个值非常高，可以解释74%的方差（0.738）。所有三个变量一起作用的贡献是显著的。

第17章

1.下面是计算卡方值的数据表。

分类	O（观察频数）	E（预期频数）	D（偏差）	$(O-E)^2$	$(O-E)^2/E$
共和党	800	800	0	0	0.00
民主党	700	800	100	10 000	12.50
无党派	900	800	100	10 000	12.50
卡方					25.00

自由度为2、显著水平为0.05的情况下拒绝零假设需要的临界值是5.99。实际值是25，表明应拒绝零假设，我们得出的结论是不同党派的投票人数具有显著性差异。

2.下面是计算卡方值的数据表。

分类	O(观察频数)	E(预期频数)	D(偏差)	$(O-E)^2$	$(O-E)^2/E$
男孩	45	50	5	25	0.5
女孩	55	50	5	25	0.5
卡方					1.0

自由度为 1、显著水平为 0.01 的情况下拒绝零假设需要的临界值是 6.64。实际值是 1,表明我们不能拒绝零假设,也就是踢足球的男孩和女孩的数量没有差异。

3.问题 3b 和 3c 适合进行卡方检验,因为收集的数据属于定类数据(都是通过/没通过)。问题 3a 和 3d 处理的数据是连续数据(分别是平均成绩和跑步速度)。

4.如图 17.6 中所见,新项目是有效果的,因为监测血糖的病人中成功控制和不成功控制饮食者,数量上没有差异的可能性非常小(大约 2.6%)。

函数库

D9 =CHISQ.TEST(C2:D3,C6:D7)

	A	B	C	D	E	F
1	实际值		及格	不及格		
2		参与课外活动	22	18		
3		没有参与课外活动	18	22		
4						
5	预测值		及格	不及格		
6		参与课外活动	20	20		
7		没有参与课外活动	20	20		
8						
9				0.371093		
10						
11						

图 17.6 使用 CHISQ.TEST 函数计算概率值

附录E　学习回报——布朗尼配方

在统计学课本中为什么介绍布朗尼配方？好问题。

严肃地说，你们肯定很努力地学习，不论是为了这门课程的分数，还是仅仅为了提高自己。你们的努力应该有所回报，这个就是。这个配方基于几个不同的配方进行了一些调整，它属于本人，很高兴与你们分享。我要把配方的全部秘密都告诉你们。

布朗尼刚从盘子里拿出来，即使没有冷却，与冰激凌搭配也非常美味。放的时间长一点就会更好吃，有嚼劲儿。刚从冰箱里拿出来就很好吃。如果冷冻，你需要消耗一些卡路里来解冻咬在嘴里的布朗尼，这比布朗尼自身的卡路里要高，这样就会有净卡路里消耗。你想吃多少冷冻的卡路里就吃多少。

1 块黄油（八勺）

半勺盐

1 杯面粉

1 勺香草

6 盎司巧克力屑（或更多）

4 盎司不加糖的巧克力（或者更多）

2 个鸡蛋

2 杯糖

2 勺蛋黄酱

1 整杯核桃

怎么做呢？

1.烤箱预热到 325℉ *。

2.在一个平底深锅里融化巧克力和黄油。

3.在一个碗里将面粉和盐搅拌在一起。

4.将糖、香草、核桃、蛋黄酱和鸡蛋加入巧克力和黄油已经融化了的锅里，搅拌。

5.第 4 步搅拌完成后倒入面粉里，搅拌。

6.加入巧克力碎屑。

7.倒入 8 英寸 × 8 英寸的抹了油的烤盘。

8.烤 35~40 分钟，或者直到牙签或筷子插进去再拿出来之后还是干净的。

注意事项

• 我知道使用蛋黄酱的争议。如果你认为不好，就不要加蛋黄酱。这些布朗尼不会因为任何配料就不好吃了，是否加蛋黄酱就看你自己的喜好了。

• 使用好的巧克力——可可脂的含量越高越好。你可以用到 6 盎司甚至更多的巧克力碎屑。

---

* 325℉（华氏度）约为 163℃（摄氏度）。







01 如何撰写研究计划书

Proposal，这个词是研究生们熟悉的“开题报告”，也是申请科研项目经费资助时必须填写的“课题论证书”或“课题申请书”，还有学者称之为“提案”，本书翻译为“研究计划书”。

本书分为三大部分：第一，撰写计划书；第二，申请研究资助；第三，计划书样例。

330 页的篇幅，意味着丰富的内容、详细的细节。





02 会读才会写：导向论文写作的文献阅读技巧

独创的阅读密码表，让读者在阅读学术论文时，能够进行系统化的阅读、摘记，并以易辨识、易提取的格式实现海量信息的组织，最终导向有效率的论文写作。





03 APA 格式：国际社会科学学术写作规范手册

APA 格式是国际社会科学学术杂志主流使用的学术论文写作和投稿规范。

从 1929 年第 1 版起，这套规范不断丰富、完善，第 6 版可谓一部名副其实的世纪经典。





04 管理学问卷调查研究方法

著名管理研究学者徐淑英推荐。

尽管书名指向“管理学”，但只是书中的例子较多来自管理学，实则社会科学皆通用。

全面阐释了问卷调查研究所需要的几乎所有方面的知识和技能，包括文献回顾、理论发展、假设检验，以及目前被最广泛运用的问卷数据统计分析方法。



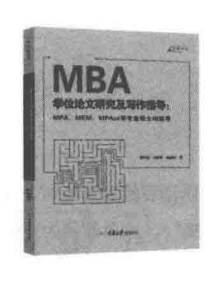


05 社会科学研究：从思维开始（第 10 版）

10 版，意味着不断完善，三十年来获得一代又一代读者认可。

引导读者一步步领会“科学”的要义，充分理解科学研究的关键所在。每一个将要做、正在做、还要做社会科学研究的人都可从中受益。

本书相关例题数据和练习题数据集，可加万卷方法读友（QQ）群 600601451 获得。





MBA 学位论文研究及写作指导

为 MBA、MPA、MEM、MPAcc 等管理类专业学位的硕士生量身打造的论文指导书。

还原研究论文与工作职业之间的关系，坚持求学应以提升解决问题的能力为目的，遵循有工作经历就能写出好论文的指导思路。

对提出有效的研究问题所总结的技巧，经济和管理相关专业的研究生均可受用。





社会研究方法（第 2 版）

一位中国教师三十余年教授研究方法课程的经验和心血之作。

全书知识点体系，完全针对中国学生的学习和研究情境。

强调研究问题和问题意识的重要性。





教育研究：定量、定性和混合方法

本书客观地告诉大家每一种研究方法的长处及其适用情况，展示专家们如何进行高水平的研究，以及如何选择合适的研究方法。每一种方法的介绍都同样详细，如能仔细而正确地使用，书里所有的研究方法都能在你的研究中发挥其优点。





顺利完成硕博论文：关于内容和过程的贴心指导

一部被美国多所高校选为论文写作教材的“名著”。

两位资深博导，对研究生面临的困难、犯错的“雷区”，有真切的体会和丰富的经验。

这些，都转化为书中细致入微的指导意见。

不仅仅提供研究方法和技术的指导，还针对论文写作中经常发生的情绪障碍、任务障碍、人际困难，给予真诚建议。

“写作”那一章非常棒，特别适合从小就讨厌写作文的人。





爱上统计学

《爱上统计学：excel》的姊妹篇，所使用的软件是SPSS。本书清晰地阐明了整个抽样调查、统计检验的思想和逻辑。书中总结的统计检验路径选择表，对搞不清什么时候用独立样本T检验，什么时候用卡方检验的入门新手特别有用。